COST-EFFECTIVE MANAGEMENT OF CHRONIC DISEASES: SURVEILLANCE, TREATMENT, AND ELIMINATION

A Thesis
Presented to
The Academic Faculty

by

Qiushi Chen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2016

COST-EFFECTIVE MANAGEMENT OF CHRONIC DISEASES: SURVEILLANCE, TREATMENT, AND ELIMINATION

Approved by:

Professor Turgay Ayer, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Jagpreet Chhatwal
Harvard Medical School
Massachusetts General Hospital

Professor Pinar Keskinocak
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Ton Dieker
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Paul Griffin
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: November 7, 2016

To Chang and Ying, my parents.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor Prof. Turgay Ayer for his absolute support of my PhD study. His guidance has helped me in every aspects of being a good researcher: how to discover problem, mature research idea, write up manuscript, and disseminate work. His motivation, patience, and positive attitude have helped me through many difficult periods during my research. He has set an excellent career model for me to follow.

My sincere thanks also go to my co-advisor Prof. Jagpreet Chhatwal at Harvard Medical School for his help, support, and advice without reservation over years. Thanks to many opportunities that he has provided for me, I am fortunate to work on many important research problems with high practical impact, to collaborate with many brilliant domain experts, and to work at different environments as my summer research interns, which altogether help to widen my vision in research.

Besides my advisor, I would like to thank the rest of my thesis committee, Prof. Pinar Keskinocak, Prof. Ton Dieker, and Prof. Paul Griffin, for taking time out from their busy schedule to serve on my thesis committee and providing their insightful comments and encouragement. I also sincerely appreciate their generous help and advice during my academic job searching process.

Very special thanks go out to Dr. Christopher Flowers in Emory University, my first mentor in clinical research. He has provided enormous help when I first started interdisciplinary research across engineering and medicine. He is always supportive, constructive, and patient, and guides me to understand the clinical context of research problems. I must also acknowledge Dr. Daniel Goldstein, Dr. Nitin Jain, and other extraordinary clinical experts and researchers in Winship Cancer Institute of

Emory University with whom we have been collaborated in various projects. Their motivation, domain expertise, and invaluable insights are the prerequisites to many of our successful studies.

I am grateful to my fellow students and friends accompanying me in the past years. Many thanks to Can Zhang, Jan Vlanchy, Anthony Bonifonte, Caglar Caglayan, Jia Yan, and Zhaowei She for being an excellent research group together, and I have enjoyed and benefited much from our discussion on various topics including but not limited to research. I am also fortunate to have friends Xian Wang, Fan Ye, Wenjia Wang, Fang Cao, Zhihao Ding, Qianyi Wang, Weijun Xie, Yijun Han for sharing the fun as well as difficult time that I have been through over the past years. My life will certainly be less colorful without them. I am also grateful to Dr. Lei Zhao in Tsinghua University as my first academic advisor during my undergraduate study for enlightening me the first glance of research.

Last but not the least, I would also like to thank my beloved families, my parents and grandparents, for their 24/7 unconditional support and care through my entire life. Without their understanding and encouragement, I would not have achieved such a potential of my career and life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Chronic diseases are the leading causes of death and disability in the United States and most of the developed world. Chronic diseases also account for the vast majority of healthcare spending, and the cost of care continues to rise. Mathematical models and operations research methods may serve as useful tools to systematically analyze chronic disease management problems, derive useful insights, and possibly inform policy decisions. This dissertation addresses three important problems related to care of cancer and hepatitis C, and uses quantitative analyses to inform the best management strategies for these chronic conditions.

We divide this thesis in four chapters. In Chapter 1, we provide overall motivation and highlight our main findings and contributions. In Chapter 2, we study the economic and disease burden of chronic lymphocytic leukemia (CLL) with the emerging therapeutic options. Recently available oral targeted therapies represent a significant advance for the treatment of CLL, but their high cost has raised concerns about affordability and value to the society. To evaluate the impact of new oral therapies for CLL on the economic and disease burden, we develop a comprehensive simulation model to project the economic and clinical outcomes for the CLL population with the evolving treatment strategies. Our results show that the oral targeted therapies will substantially increase the annual cost of CLL management and per-patient lifetime out-of-pocket cost by more than 5-fold from the year 2011 to 2025, which far outpaces the rising cost of treating other cancers. At current prices, the new oral targeted treatment strategy is deemed not cost-effective at the willingness to pay threshold of $100,000/quality-adjusted life-year. Our findings remain robust against parameter uncertainties through extensive sensitivity analyses. Our results highlight that such

an economic impact could result in financial toxicity, limited access, and lower adherence to the oral therapies, which may undermine their clinical effectiveness. More sustainable pricing strategy for oral targeted therapies is imperatively needed.

In Chapter 3, we study optimal surveillance of liver cancer in hepatitis C-infected population. In the United States, hepatitis C-related liver cancer has become the fastest-growing cause of cancer-related deaths. Although surveillance in hepatitis C patients has shown improved early detection of cancer, the optimal use of surveillance remains unknown. In this study, we develop a mixed-integer programming (MIP)-based framework to systematically examine the cost-effectiveness of different surveillance policies and identify the optimal policy that will maximize societal net benefit. Our MIP-based framework captures two problem features that make dynamic programming-based formulation computationally intractable. In particular, our proposed framework allows to (1) explicitly formulate a class of policies that are practical for implementation, which we call "M-switch policies", and (2) tailor surveillance policies for each subpopulation by stratifying surveillance intervals based on the observable disease states. We analyze structural properties and a) identify the sufficient conditions under which additional surveillance improves the cost-effectiveness of a policy, b) characterize when the surveillance policies should be adapted to populations with different disease progression rates, and c) quantify the trade-off between decreasing cancer incidence and increasing treatment outcomes. We parameterize our model using data from a clinical trial, a previously validated simulation model, and published clinical studies. Our numerical analyses provide two main results with important policy implications. First, unlike the current one-size-fits-all type policies, the optimal surveillance interval should be stratified based on the stage of hepatitis C infection and age; second, expanding surveillance to patients in earlier stage of hepatitis C infection improves the cost-effectiveness of liver cancer surveillance program.

In Chapter 4, we study an optimal resource allocation problem motivated by

the ongoing efforts by the World Health Organization to eliminate hepatitis C virus (HCV) by 2030. More than 170 million people are chronically infected with HCV globally. Although new antiviral treatments for HCV offer a hope to eliminate HCV, most countries do not have national programs to screen and treat HCV. To guide the policy decisions in hepatitis C epidemic control, we develop two optimal control formulations considering HCV screening and treatment interventions. In the first formulation, we consider a cost minimization problem subject to a final target prevalence constraint. We analytically show that the optimal treatment is a pure bang-bang policy without singular arcs for any given screening rate. The simple policy structure with constant screening and treatment rates does not result in a significant loss compared with policies with dynamic rates, which represents a simple, effective and practical policy structure for implementation. In the second formulation, we consider a budget allocation problem that aims to minimize the disease burden subject to a fixed budget. We show that optimal allocation policy follows a simple treatment-first rule both analytically and numerically. Using this model framework, we numerically solve the optimal control problems and identify optimal screening and treatment strategies for the case of HCV elimination in India.

CHAPTER I

INTRODUCTION

Chronic diseases and conditions, such as cancer, cardiovascular disease, and arthritis, have become the most common health problems worldwide. Chronic diseases are the leading causes of death since the end of twentieth century. They contributed to more than 62% of all deaths worldwide in 2009, and this number is expected to increase to 73% by 2020 (88). Chronic diseases are also prevalent in the United States (US). Data from 2005 showed that 45% Americans—133 million people—had at least one chronic condition, and the prevalence is predicted to increase to 157 million by 2020 (233). Chronic diseases are responsible for 7 out of every 10 deaths of Americans each year; in particular, cardiovascular disease and cancer together account for 46% of all deaths (234).

Chronic diseases are costly. Management of chronic diseases accounts for the vast majority of health care spending in the US. Eighty-six percent of all US health care spending, about $2.2 trillion, was attributed to chronic disease in 2010 (83), including 81% of all hospital admissions and 76% of all physician visits (12). Moreover, the cost of chronic diseases is expected to increase to $4.2 trillion by 2023 (57). Such a substantial amount of spending on chronic disease is not proportional to the percentage of people with chronic diseases. In particular, more than 80% of spending of private insurance, Medicare, and Medicaid is attributed to only less than 50% people with chronic diseases (12). Thus, people with chronic diseases represent the heaviest users of health care services, and management for these patients leads to substantial economic burden to society.

One of the important reasons for significant cost burden of chronic disease management is the high drug prices, especially in cancer care. Cancer drug prices have increased substantially in the past few decades (Figure 1.1). The average annual cost of cancer treatment was below $10,000 before 2000, and increased to more than $100,000 per year in 2012 (114, 136). The cost for almost all new cancer drugs approved in 2014 range between $120,000 and $170,000 per year per person (48).

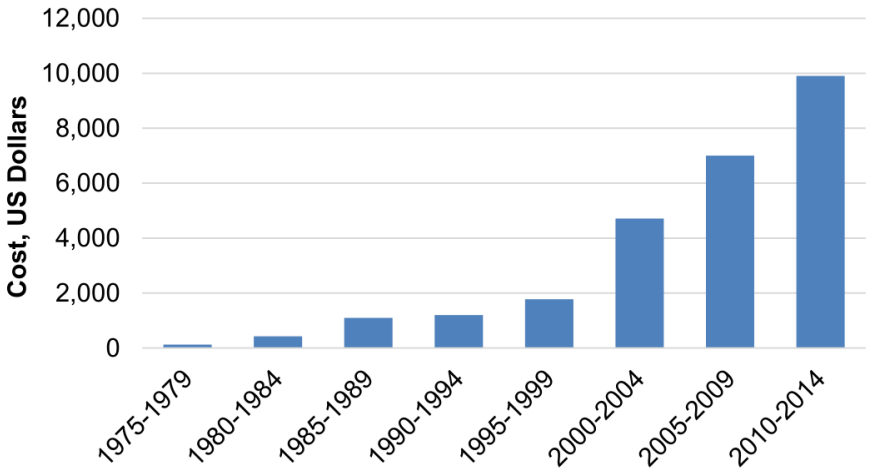


Figure 1.1: Rising cost of cancer drugs in the United States.

The median monthly cost for new cancer drugs in the United States has soared since the 1970s despite an increasing number of available brands. Adapted from (186): Saltz, L. B. (2016). "Perspectives on Cost and Value in Cancer Care". *JAMA Oncology*, 2(1), 19-21.

At such a high price, cancer treatment is hardly affordable, even for well-insured patients. Studies have shown that cancer patients were three times more likely to experience bankruptcy than those without cancer (176); 1 in 5 cancer survivors reported financial hardship associated with cancer care, and 12% could not cover the share of medical care cost (235). The financial burdens to cancer patients could limit their access to therapeutic advancements, and lower their adherence to prescribed treatment, which eventually impairs patients' well-being. Such consequence of high drug cost has recently been considered as a side effect of cancer treatment, referred

to as "financial toxicity" (222). Given the financial strain that both society and individual patients are facing, a better understanding of the financial impact and cost-effectiveness of cancer treatment options is important for effectively utilizing scarce health care resources.

The substantial economic burden associated with chronic diseases has also been fueled by the increasing incidence and prevalence, mainly due to a rise in disease-specific risk factors and an aging population (27). Since many risk factors are behavioral, such as lack of physical activities, smoking, and excessive alcohol use, which are controllable through health promotions, several chronic diseases are indeed preventable (42). In addition, early intervention can also be effective to control, or even to cure the disease at an early stage. For example, many cancers including breast, prostate, liver cancers are curable if detected in early stages through regular screenings; risks of liver cancer can even be completely eliminated in hepatitis C infected patients if their hepatitis is cured in early stages. As a result, health intervention policies can play a critical role in disease prevention and control, as well as in alleviating the stress of total disease and economic burden on the society.

To better support health policy decisions in chronic disease management, mathematical models can serve as useful tools. Because of time and resource constraints, randomized controlled trials, the gold standard in medicine, are not suitable for assessing and comparing various policies for chronic disease management at the population level. Alternatively, mathematical models may be useful in many ways. First, models are helpful in evaluating health economic outcomes of novel interventions and strategies that have not yet become the standard of practice. For instance, models can project the economic impact of emerging therapeutic options, and assess the cost-effectiveness of multiple possible treatment options. Second, mathematical models can be used to optimize complex decisions with respect to complex dynamics of the

disease in the population, where all alternative options cannot be exhaustively enumerated and tested through real-world experiments, such as sequential decisions of cancer screening, and dynamic resource allocations in epidemic control.

Motivated by these research opportunities, we focus on modeling and analysis of three important questions related to the care for cancer patients and elimination of hepatitis C in this dissertation. In Chapter 2, we study the economic and disease burden of chronic lymphocytic leukemia (CLL) in the era of oral targeted therapies in the US. Novel oral targeted agents have shown marked improvement in patients' survival, but their high cost has raised concerns about their affordability and the economic impact to payers, patients, and society. We develop a comprehensive simulation model to evaluate economic and clinical outcomes for the CLL population. This model captures the evolving treatment strategies with emerging oral therapies, and dynamics of disease incidences in an aging population. Our analyses show that oral targeted therapies will substantially increase the cost of CLL management from both payers' and patients' perspective. The increasing cost burden is driven by high drug prices, improved survival, and extended treatment duration. At current prices, these therapies would not be considered cost-effective to become the new standard of care.

In Chapter 3, we study the liver cancer surveillance problem for the hepatitis C-infected population. In the US, hepatitis C-related liver cancer has become the fastest growing cause of cancer-related death. Although surveillance in hepatitis C-infected patients has shown improved early cancer detection and survival outcomes, the optimal surveillance policy remains unknown. In this chapter, we develop a mixed-integer programming-based framework to identify the most cost-effective surveillance policies. Our proposed framework allows (1) explicit formulation of "M-switch" policy structures that are practical for implementation, and (2) surveillance policies tailored for subpopulations which are stratified by observable clinical factors. We find that,

unlike current one-size-fits-all policies, the optimal surveillance interval should depend on patients' liver fibrosis stage and age. Moreover, expanding surveillance to patients with earlier fibrosis stage improves overall cost-effectiveness of liver cancer surveillance program.

In Chapter 4, we study an optimal resource allocation problem motivated by the ongoing efforts by the World Health Organization to eliminate hepatitis C virus (HCV) by 2030. Global HCV burden is massive: more than 170 million people are chronically infected with HCV. Although elimination is theoretically feasible with therapeutic and diagnostic tools that are currently available, the challenges persist due to ongoing transmission, unawareness of infection, high treatment cost, and limited resources. Most countries do not have any national program to screen and treat HCV. Several questions need to be answered such as how to allocate resources between screening and treatment to reach HCV elimination targets. In this study, we consider two different intervention modes, namely *screening* and *treatment* for HCV. We analyze the optimal intervention policies in two problem formulations, and showcase the numerical solutions and policy implications in a case study of hepatitis C elimination in India. In the first formulation, we consider a cost minimization problem subject to a final target prevalence constraint. We analytically show that the optimal treatment is a pure bang-bang policy without singular arcs for any given screening rate. The simple policy structure with constant screening and treatment rates does not result in a significant loss compared with policies with dynamic rates, implying a simple, effective and practical policy for implementation. In the second formulation, we consider a budget allocation problem that aims to minimize the disease burden subject to a fixed budget. We show that optimal allocation policy follows a simple treatment-first rule in both analytical and numerical results.

CHAPTER II

PREVALENCE AND ECONOMIC BURDEN OF CHRONIC LYMPHOCYTIC LEUKEMIA

2.1 Background

Chronic lymphocytic leukemia (CLL) is the most prevalent leukemia in the western world. In the US, there are approximately 130,000 patients living with CLL, and about 15,000 new cases occur every year (200). While most patients with CLL have early stage disease at the time of initial diagnosis and are recommended for watchful waiting (107), the majority eventually require treatment, typically after a few years of observation and often experience prolonged survival (92, 96).

Chemoimmunotherapy (CIT) regimens such as fludarabine, cyclophosphamide, and rituximab (FCR) have been the standard first-line treatment for young patients with CLL (219, 95). In a single-center experience, FCR regimen led to a complete remission (CR) rate of 72% and median progression-free survival (PFS) of 80 months (115, 212). Subsequently, the German CLL8 trial established FCR as the standard first-line therapy (95). For patients with age >65 years or with comorbidities, the combination of chlorambucil and obinutuzumab is considered standard of care, based on the results of the CLL11 trial (84).

In the last few years, major strides have been made in understanding the biology of CLL and this led to significant advances in the treatment of CLL. In particular, oral targeted agents such as ibrutinib and idelalisib have demonstrated remarkable outcomes in patients with CLL. In the relapse setting, ibrutinib showed an overall response rate (ORR) of 90% and an estimated PFS of 69% at 30-months (38); idelalisib showed an ORR of 72% and a median PFS of 15.8 months as monotherapy

in a phase I study (32), and an ORR of 81% and estimated overall survival (OS) of 92% at 12 months when used in combination with rituximab in a phase III study (81). Subsequently, in 2014, the Food and Drug Administration (FDA) approved two oral targeted therapies: ibrutinib for patients with relapsed/refractory CLL and for patients with del(17p), and idelalisib in combination with rituximab for patients with relapsed/refractory CLL. In March 2016, ibrutinib was approved for first-line management of CLL. In addition, several other targeted therapies are expected to become available in the near future (108). Venetoclax was approved for patients with relapsed CLL with del(17p) in April 2016 (77). These novel therapies have revolutionized the CLL treatment paradigm.

However, the high cost of these targeted therapies raises concerns for payers as well as patients (192, 235). Both ibrutinib and idelalisib are priced around $130,000 per year and are recommended to be taken until patients have progressive disease or significant toxicities. In contrast, the costs for CIT-based treatments range from $60,000-100,000 for a finite duration, a typical 6-cycle course lasting for about 6 months. Therefore, novel targeted therapies could strain the budget of both private and government payers, such as Medicare, and co-payments and other expenses can be a substantial burden to patients. Yet, the budget impact and cost-effectiveness of these therapies are not well understood. In this study, our objective was to project the changing economic as well as disease burden of CLL in the US in the era of targeted therapies, and to evaluate the affordability and value of these new therapies.

2.2 Methods

We developed a microsimulation model, simCLL (simulation model of CLL management), that simulated the dynamics of patient population under given management strategies in the US from 2011 to 2025.

2.2.1 Patient population characteristics

Patient characteristics were defined by age, phase of CLL treatment (watchful waiting, first-line, or relapse), and del(17p) status. The age at diagnosis of each individual patient was sampled from the age distribution based on the SEER data between 2000-2011. (211). Del(17p) was assumed to be present in 7% of CLL patients (95). New CLL cases were added to the simulated population in each year based on the published annual incidence estimates from the American Cancer Society (200) (Table A.1). The prevalence of CLL from the simulation model was calibrated to Surveillance, Epidemiology, and End Results (SEER) data between 2000 and 2011 (Figures A.1 and A.2).

2.2.2 Simulated clinical pathways

We modeled the clinical course of patients with CLL using a patient-level state-transition model, which included the following health states: watchful waiting, first-line treatment, relapse, and death (Figure A.3). Majority of newly diagnosed patients do not need immediate treatment (96) and were assumed to start in the watchful waiting state (229) (the probability was determined through the model calibration). Upon the failure of the first-line treatment, patients entered relapse state. The probabilities of health state transitions were estimated based on time-to-treatment, progression-free survival, and overall survival data observed in clinical trials (Tables 2.1, A.2, and A.3). We selected the trials representing the best available evidence (e.g., phase III trial, or large observational studies) for the major regimens in general practice, with the reference to clinical guidelines (96) and expert opinions. We also validated our model by comparing the simulated survival curves with observed survival data (Figure A.4). Finally, we applied age-specific background mortality based on the US life tables (16).

Table 2.1: Summary of treatment parameters

Treatment	PFS/OS	CR/PR rate (%)	Adverse events (%) (Anemia/Neutropenia /thrombocytopenia /infection)	Drug cost ($ per cycle)	Administration cost ($)
First-line setting					
FCR (95)					
Fit patients	Median PFS=51.8 mo 3-yr PFS=65% 3-yr OS=87%	44/46	5/34/7/25	7455 (cycle 1) 10063 (cycle 2-6)	817 (cycle 1) 716 (cycle 2-6)
Del(17p) patients	Median PFS=11.3 mo HR=7.49 for PFS HR=9.32 for OS	5/63	5/34/7/25	7455 (cycle 1) 10063 (cycle 2-6)	817 (cycle 1) 716 (cycle 2-6)
GClb (unfit patients) (84)	Median PFS=26.7 mo	20.7/57.7	4/33/10/12	19063 (cycle 1) 6679 (cycle 2-6)	882 (cycle 1) 221 (cycle 2-6)
Clb (unfit patients)[a] (67)	Median PFS=18 mo	0/51	27/12/20/4	779	-
Ibrutinib					
Fit/unfit patients [b] (36)	18-mo PFS=90% 24-mo OS=98%	4/82	6/10/2/6[c]	10270 (for 4 wks)	-
Del(17p) patients (72)	24-mo PFS=91% 24-mo OS=84%	12/85	14/24/10/6	10270 (for 4 wks)	-
Relapse setting					
FCR (230)	Median OS=42 mo	-	24/81/34/16	7455 (cycle 1) 10063 (cycle 2-6)	817 (cycle 1) 716 (cycle 2-6)
BR (76)	Median OS=33.9 mo	-	16.6/23.1/28.2/12.8	13619 (cycle 1) 16227 (cycle 2-6)	419

Table 2.2: Summary of treatment parameters (continued)

Treatment	PFS/OS	CR/PR rate (%)	Adverse events (%) (Anemia/Neutropenia /thrombocytopenia /infection)	Drug cost ($ per cycle)	Administration cost ($)
Relapse setting					
Ofatumumab[d] (231)	Median OS=13.7 mo d	-	-/14/-/12	36008 (cycle 1) 18290 (cycle 2-6)	1108 (cycle 1) 398 (cycle 2-6)
Idelalisib+R[e] (81)	12mo-OS=92%	-	5/34/10/-	Rituximab: 23469 (cycle 1) 10865 (cycle 2-5); Idelalisib: 8862 (for 4 wks)	662 (cycle 1) 276 (cycle 2-5)
Ibrutinib (38)	30mo-OS=79%	-	-/18/10/51	10270 (for 4 wks)	-
HSCT	5yr-OS=51% 55	-		217573 for first 100-day	-

FCR, fludarabine, cyclophosphamide, and rituximab; Clb, chlorambucil; GClb, obinutuzumab plus chlorambucil; Idel, idelalisib; R, rituximab; BR, bendamustine plus rituximab; HSCT, hematopoietic stem cell transplantation; PFS, progression-free survival; OS, overall survival; CR, complete response; PR, partial response; TN, treatment naive; mo, month.

Note. a: For patients with del(17p), PFS and OS outcomes for Clb treatment were extrapolated by applying the hazard ratio associated with del(17p) observed in CLL8 trial (95) to the survival estimates for unfit patients, and no CR/PR were observed from Eichhorst et al. (2009)(67).
b: Assume that fit patients have the same survival outcomes as unfit patients with first-line ibrutinib, given no direct follow-up data have existed for fit patients in the first-line setting.
c: 6% risk of atrial fibrillation has been observed in ibrutinib arm (36).
d: We used the data of FA-ref arm in that study.
e: Do not differentiate del(17p) patients as the data were not stratified by del(17p) status.

2.2.3 Treatment strategies

A treatment strategy defined the specific therapy for a patient by his status of relapse, fitness (determined by age), del(17p), and the year of treatment (Figure 2.1). We first simulated a clinical scenario that considered the current standard-of-care and emerging treatment options (Figure 2.1A). In particular, prior to 2014, CIT was the mainstay treatment for patients with CLL. The most common choices for the first-line treatments were FCR for fit patients and chlorambucil for unfit patients. From 2014 onwards, the oral targeted therapies were approved for patients with relapsed CLL and for patients with del(17p) (37, 81). From 2016 onwards, oral targeted therapies became the standard-of-care in the first-line setting (36). The scenario described above is referred to as the *oral targeted therapy scenario*.

For comparison, we simulated a scenario where CIT would have remained the standard-of-care in future (Figure 2.1B). In this scenario, first-line unfit patients would receive obinutuzumab plus chlorambucil after 2014 (84). This scenario is referred as *chemoimmunotherapy scenario*.

2.2.4 Costs

Direct medical costs were considered, including the cost of drugs and administration, the cost of routine follow-up, and the cost of management of adverse events. Drug costs were calculated based on the doses of the standard regimen and the average sales price (ASP) of each drug. ASP was estimated as 26% lower than the average wholesale price (AWP (2), see estimates in Tables 2.1-2.2, and details in Table A.5) as suggested by an Office of Inspector General study (134). For oral targeted agents, drug costs were accumulated for an indefinite period until treatment was discontinued. As observed in the clinical studies with ibrutinib, 87% of patients in first-line setting and 75% in relapsed setting would continue oral targeted therapy beyond 18 months (36, 146).

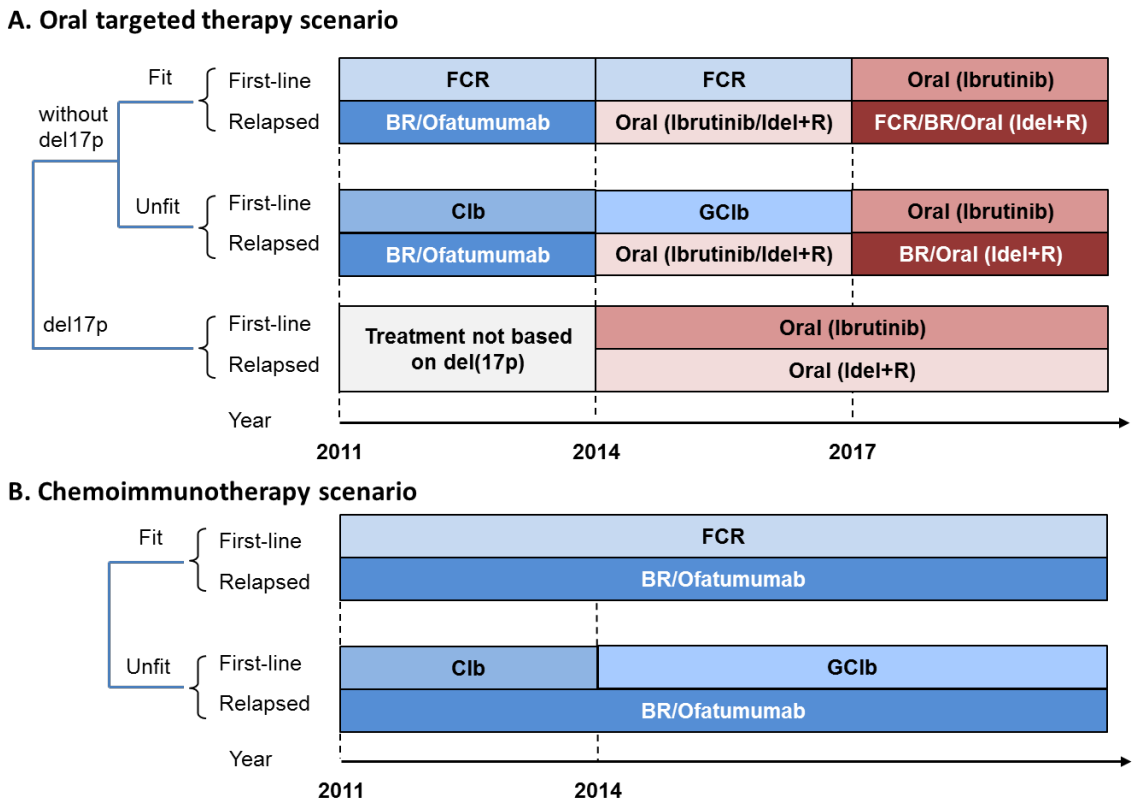


Figure 2.1: Management strategies for chronic lymphocytic leukemia (CLL) patients. (A) The oral targeted therapy scenario with evolving therapeutic options for CLL patients. (B) The chemoimmunotherapy scenario which continues to use chemoimmunotherapy as the standard of care. We assumed equal allocation to multiple therapies if more than one therapies are considered for patients in the same condition. For example, for fit patients in the relapse setting during 2014–2017, 50% patients receive ibrutinib and 50% receive idelalisib plus rituximab. Moreover, 5% relapsed patients were assumed to receive hematopoietic stem cell transplantation (now shown in figure). FCR, fludarabine, cyclophosphamide, and rituximab; Clb, chlorambucil; GClb, obinutuzumab plus chlorambucil; Idel, idelalisib; R, rituximab; BR, bendamustine plus rituximab.

Administration costs such as physician visit and chemotherapy infusions were calculated based on the Medicare physician fee schedule (43) (Tables 2.3 and A.4, using the approach described elsewhere (221). Common serious adverse events, including grade 3/4 infection and hematological toxicities, were simulated based on the reported incidence for each therapy (Tables 2.1-2.2), and the management cost for each was derived from the corresponding treatment and hospitalization costs (Table 2.3) (44). We also considered the risk and the cost of atrial fibrillation with oral targeted therapy because it has emerged as an important concern with ibrutinib (36).

Table 2.3: Model input parameters.

Variable	Base value	Range	Reference
Probabilities			
Prevalence of del17p	7%		(95)
Probability of WW at diagnosis	0.85		Calibrated
Probability of fitness			
Age: <65	0.95		Assumption
Age: 65-70 [a]	0.2		Assumption
Probability of initiating first-line treatment	Estimated from time-to-treatment		(171)
Probability of discontinuing oral targeted therapy at 18 months:			
First-line treatment [b]	0.13	(0.109, 0.151)	(36)
Relapse [c]	0.245	(0.147, 0.342)	(146)
Health utilities			
First-line			(25, 151)
Complete response	0.91		
Partial response	0.84		
No response	0.78		
Relapsed	0.68		(25, 151)
Disutility for FC treatment periods	-0.07		(4)
Cost ($)			
Chemo intravenous infusion:			
First hour	135.87		CPT96413 (43)
Additional hour	28.25		CPT96415 (43)
Each additional seq	62.93		CPT96417 (43)
Office/outpatient visit	51.13		CPT99213 (43)
Blood test	80		(93)
Cost of adverse events			
Anemia	1967	(1910, 1998)	(44)
Neutropenia	3207	(2885, 3539)	(44)
Thrombocytopenia	1136	(620, 1191)	(44)
Infection	12097	(7418, 28926)	(44)
Atrial fibrillation	17342	(16123,18322)	(131)

a: Assume all patients older than 70 at diagnosis are unfit patients;
b, Range is calculated based on the confidence interval of binomial distribution, corresponding to treatment discontinuation probabilities of 0.007 (0.006, 0.008) for each 4-week cycle;
c, Correspond to treatment discontinuation probabilities 0.014 (0.008, 0.021) for each 4-week cycle.

2.2.5 Health-related quality-of-life

The health-related quality of life weights (utilities) were adjusted by health states and patients' age (25, 75). We assumed utility of one in watchful-waiting state. For the first-line treatment state, a patient's utility was determined by his response type (i.e., complete, partial, or no response), which was sampled according to the response rate of his treatment. In addition, the utilities were adjusted based on patients' age (97), and a disutility was applied to fludarabine and cyclophosphamide-containing regimens (41, 4).

2.2.6 Model outcomes

We projected the number of people living with CLL and the annual cost of CLL management in the US from 2011 to 2025. In addition, we calculated per patient life-time cost with oral targeted therapies as well as with CIT as the standard-of-care. Because the majority of patients are older than 65 at the time of CLL diagnosis and covered by Medicare, we also estimated the life-time out-of-pocket cost of the oral targeted therapies for patients enrolled in Medicare Part D plan (Tables A.6 and A.7). We estimated total discounted person-LYs and person-QALYs as the total health outcomes at the population-level from 2011-2025, and finally estimated the incremental cost-effectiveness ratio (ICER) of oral targeted therapies in comparison with chemoimmunotherapy scenario. All costs were converted to 2015 US dollar. In the cost-effectiveness analysis, future outcomes were discounted to value in 2015 at 3% per year (64). For simplicity we presented the rounded values of all our numerical results.

2.2.7 Sensitivity analysis

To evaluate the robustness of outcomes against uncertainty in model inputs, we performed one-way deterministic sensitivity analyses. Utilities and probabilities of discontinuation of oral targeted therapies were varied within their reported confidence

interval, other transition probabilities and costs were varied within 20% range, and survival distributions were adjusted with hazard ratios between 0.8 and 1.2. We also performed probabilistic sensitivity analysis (PSA) that accounted for joint uncertainty in all model inputs.

In addition, we performed two scenario analyses. First, considering the aging US population (Table A.8) (170), we adjusted the CLL incidence with the US population projection data (Table A.9) and simulated a scenario with a higher CLL incidence rate. Second, we evaluated a scenario simulating partial uptake of oral targeted therapies representing a gradual transition from CIT. Specifically, we assumed 25% utilization of oral targeted therapies in first-line treatment for fit patients after the approval in 2016, and increased by 25% every year until a full utilization in 2019 and onwards.

2.3 Results

2.3.1 Disease burden

The total number of people living with CLL is projected to increase from 128,000 in 2011 to 199,000 (55% increase) in 2025 because of improved survival with the use of oral targeted therapies. In contrast, if CIT remained the standard-of-care, the number of people living with CLL would be 162,000 (26% increase) by 2025 (Figure 2.2).

2.3.2 Cost burden

Annual cost of CLL care. Under the *oral targeted therapy* scenario, the annual cost of CLL management is projected to increase from $0.74 billion in 2011 to $5.13 billion (593% increase) in 2025 (Figure 2.3). The first surge in the annual cost occurred in 2014 when oral targeted therapies became available for relapsed patients, and the second surge will occur in 2016 because of the approval of oral targeted therapy in

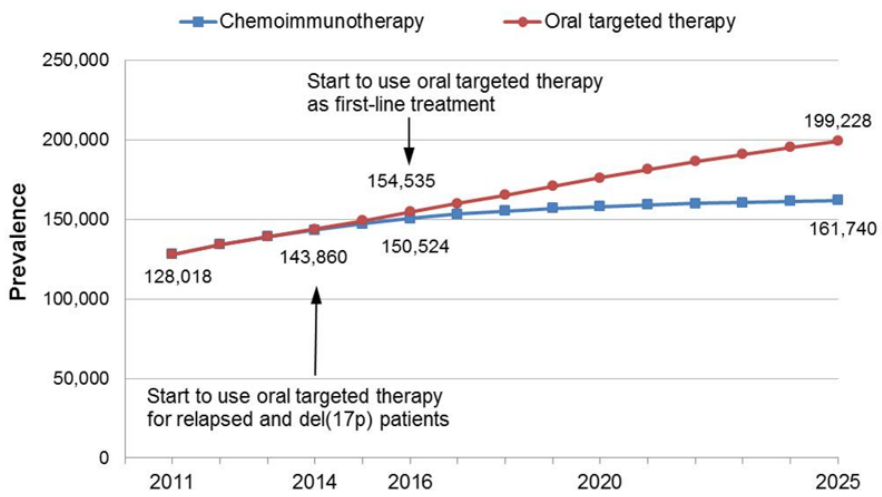


Figure 2.2: Trend in disease burden of CLL.

Note. The use of oral targeted therapies is projected to increase the number of people with CLL from 128,000 in 2011 to 199,000 (55% increase) in 2025 because of improved survival with the use of oral targeted therapies.

the first-line setting. In contrast to the increasing cost trend with oral targeted therapies, the annual cost under the *chemoimmunotherapy* scenario would have remained relatively stable from 2014 onwards, reaching $1.12 billion in 2025. Compared with chemoimmunotherapy scenario, oral targeted therapies would result in an additional spending of $29 billion from their availability in 2014 until 2025. Among the total cost of CLL management, drug costs constituted 96% in the oral targeted therapy scenario and 86% in the chemoimmunotherapy scenario.

Lifetime cost of CLL treatment. The per-person lifetime cost of CLL treatment for patients initiating therapy in 2011 was $147,000, which increased to $331,000 (125% increase) for patients initiating therapy in 2014 (Figure 2.4A). For patients initiating therapy in 2016 with oral therapies (now approved in first-line), the lifetime cost of CLL treatment is projected to reach $604,000 (310% increase from 2011).

Out-of-pocket cost for Medicare patients. The majority of patients with CLL in the US are covered by Medicare, and have drug coverage through Medicare Part D

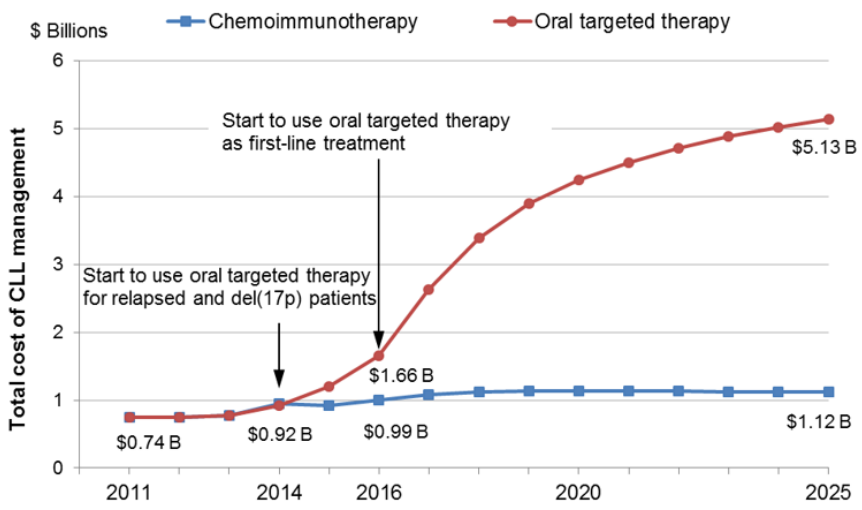


Figure 2.3: Annual management cost of CLL.

Note. The use of oral targeted therapies is projected to increase the annual cost in CLL management from $0.74 billion in 2011 to $5.13 billion (593% increase) in 2025, which is mainly driven by high drug prices, prolonged treatment duration of oral agents, and increased number of patients living with CLL.

plan. The out-of-pocket cost of oral agents for Medicare patients was estimated based on the deductible and coverage limits in the Medicare Part D plan (Table A.6), and was estimated to be $9200 for those initiating therapy in 2011, which increased to $27,000 (193% increase) for patients initiating therapy in 2014, and to $57,000 (519% increase) for patients initiating treatment from 2016 onwards (Figure 2.4B). Use of oral targeted therapies in first-line setting after 2016 also substantially increased the first-line treatment cost, which constituted the major proportion of the lifetime cost.

2.3.3 Cost-effectiveness analysis

From 2011 to 2025, the total discounted health outcomes were 1,850,000 person-QALYs (2,193,000 person-LYs) under the oral targeted therapy scenario and 1,743,000 person-QALYs (2,044,000 person-LYs) under the chemoimmunotherapy scenario. Compared with the chemoimmunotherapy scenario, the oral targeted therapy scenario resulted in an increase of 107,000 person-QALYs (149,000 person-LYs) with additional

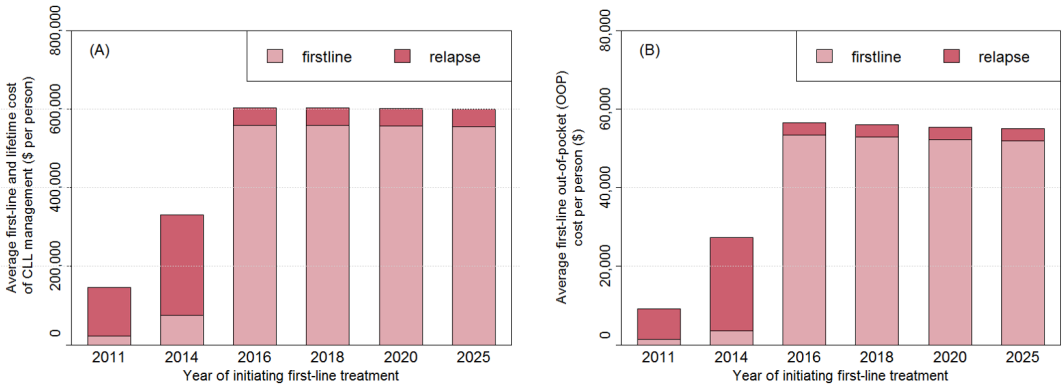


Figure 2.4: Lifetime treatment cost grouped by the year of initiating first-line treatment for the oral targeted therapy scenario.

(A) Lifetime treatment cost to payers; (B) lifetime out-of-pocket cost for Medicare patients.

discounted treatment costs of $20.2 billion. The incremental cost-effectiveness ratio (ICER) of oral targeted therapies was $189,000 per QALY ($136,000 per LY).

2.3.4 Sensitivity analyses

We examined the sensitivity of results to oral drug cost discounts. Considering a 37% discount off the AWP as the lowest price paid by private-sector payer for drug product (104), the total incremental cost was $24 billion, lifetime out-of-pocket cost for patients with first-line oral targeted therapy was $52,000, and the ICER of oral targeted therapy was $161,000/QALY (Table A.10). When the oral targeted agent cost was reduced to 50% of AWP, the ICER would reduce to $107,000 per QALY. A threshold cost analysis showed that the cost of oral targeted therapies needs to be at least 69% lower than the current AWP to bring the ICER below $50,000-per-QALY threshold.

One-way sensitivity analysis showed that the cost of CLL management was sensitive to treatment cost, discontinuation rate of oral targeted therapies, immediate treatment probability at initial visit, and time to first-line-treatment from watchful waiting state (Figure A.5). The ICER was most sensitive to survival distributions of

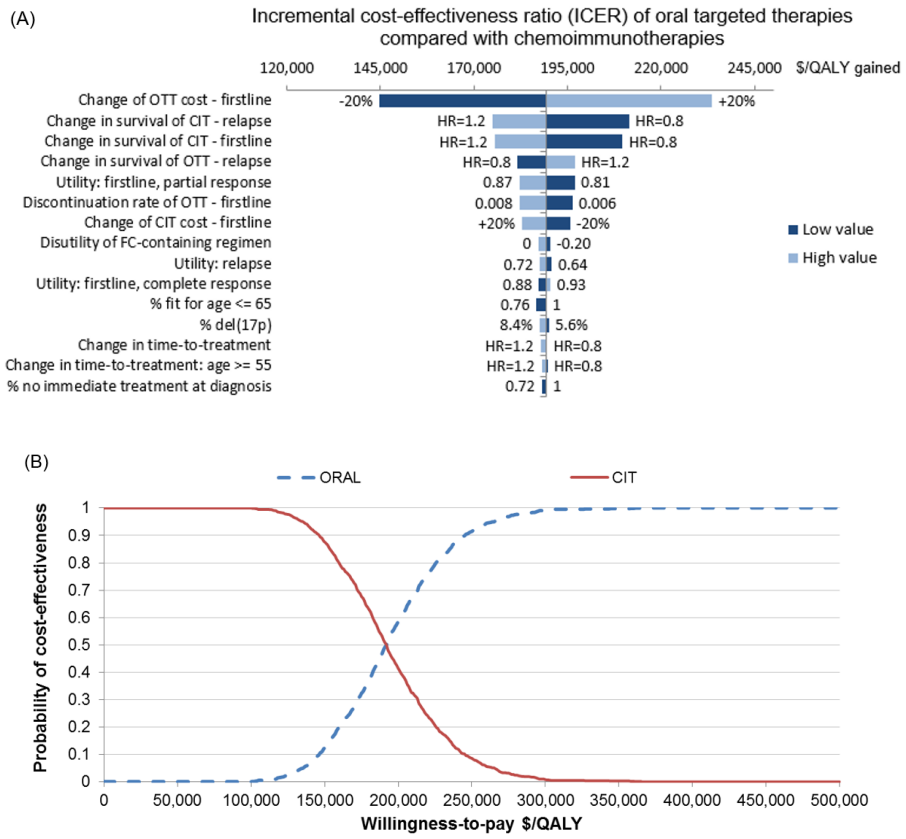


Figure 2.5: Sensitivity analysis of the cost-effectiveness of the oral targeted therapy scenario compared with chemoimmunotherapy scenario.

(A) Tornado diagram for one-way sensitivity analysis of incremental cost-effectiveness ratio; (B) cost-effectiveness acceptability curves from the probabilistic sensitivity analysis.

treatments, as well as health-related utilities for partial response in first-line treatment and in the relapse setting (Figure 2.5A). PSA showed that oral targeted therapy was deemed cost-effective with a very low probability even at a willingness-to-pay as high as \$150,000/QALY (Figure 2.5B). We also found that increase in CLL incidence because of the aging population in the US would further escalate the cost burden and reduce the cost-effectiveness of oral targeted therapies, and the partial uptake of oral targeted therapies for first-line fit patients had limited changes to all model results and would not change the conclusions (Table A.11).

2.4 Discussions

Oral targeted therapies represent a major advance for patients with CLL with improvement in overall survival compared to conventional therapies (36). Our study projected an increase in the number people living with CLL over time, largely due to improved survival in the era of oral targeted therapies. We also projected a substantial increase in the cost of CLL management in the era of targeted therapies. The annual cost of CLL management is projected to reach $5.13 billion by 2025, a 590% increase from that in 2011. The cost of new therapies will add considerable financial burden to both patients and payers. At the current price, oral therapies are not deemed cost-effective using the willingness to pay threshold of $100,000-per-QALY.

This study provides a comprehensive view and analysis of the changing burden of CLL care in the US To our knowledge, no study has evaluated the cost-effectiveness of oral targeted cancer therapies from a population level. Earlier studies on the cost of CLL did not consider recent data and changing population dynamics. One study examined the life-time cost of CLL treatment for Medicare patients equal to $87,000 using data from older drugs regimens from 1999—2007 (123). Another study by Shanafelt and colleagues estimated the annual societal cost of CLL equal to $0.73 billion using chemoimmunotherapy and $2.63 billion and $1.24 billion using ibrutinib in first- and second-line, respectively (192). However, their estimates were lower than our projections because they did not account for the growing disease population due to improved survival. Our results highlighted the expected societal impact of the rising disease burden from CLL that compounded the increased cost associated with a long-term oral therapy.

Our study has several limitations. First, our model did not consider all possible treatment sequences in practice. We also did not account for individual practice patterns that deviate from standard-of-care and guidelines because no data exist for comprehensive utilization estimates for each treatment option. However, we believe that

our approach is sufficient to capture the most commonly accepted practice patterns and population-level trends in costs and prevalence of CLL. Second, we considered constant drug prices and did not capture the possible fluctuation of drug prices over time in reality. We performed a series of sensitivity analyses on drug prices and found that our findings remained valid across a wide range of drug price discount, and that the oral therapies could be deemed cost-effective if the prices were at least 69% lower than our current AWP estimates.

Although the cost of cancer care is rising, our results indicate that the rising trend in the cost CLL management will outpace that of other cancers. The annual cost of cancer care in the US due to the aging population is expected to increase by 27∼50% from $143 billion in 2010 to $180 billion in 2020 (150). For breast and prostate cancers, the annual cost of care is expected to increase by 24∼38% from 2010 to 2020. In contrast, the annual cost of the CLL is estimated to increase by 500% from $0.7 billion in 2011 to $4.2 billion in 2020. The substantial increase in the cost of CLL management is mainly driven by the high cost of oral targeted drugs and prolonged treatment duration along with the improved survival. While our analysis suggests that the cost of CLL management will rise faster than that of other cancers, future advances in treatments could increase the costs of care of other cancers as well. Such increase could strain the budget of private as well as government payers.

Patients also will suffer from the escalating cost burden of expensive treatments, as the higher overall cost could translate into higher health insurance premium and cost sharing for individual patients (213). One study found that medical bankruptcies ranked number one constituting of 67% of all US family bankruptcies, for whom the out-of-pocket cost ranged from $18,000 to $27,000 (103). Our results showed that the lifetime out-of-pocket costs of CLL treatment for Medicare patients is expected to increase by nearly 4-fold to $57,000 for those initiating first-line oral targeted therapy after 2016, which could further exacerbate the likelihood of medical bankruptcy

and result in discontinuation of treatment. Especially for patients from low-income families, the high out-of-pocket cost of oral targeted therapies could result in limited access to these therapies and thus adversely affecting their clinical outcomes. In addition to the material financial hardship—a survey showed 12% of cancer patients could not cover their share of medical care costs, the high out-of-pocket costs could also lead to psychological financial hardship (235). Furthermore, high out-of-pocket costs could result in disparities in access to these therapies. For instance, CLL patients with lower income level may not be able to afford these therapies, adversely affecting their outcomes. Their health could remain suboptimal even in the era of oral targeted therapies (111, 177).

The high drug price has been a disturbing concern not only in the area of CLL management, but also in the setting of cancer care in general. The cost of cancer care has been rising drastically in the past 15 years. The average annual cost of cancer treatment before 2000 was below $10,000, which has now increased to more than $100,000 (114, 136). A recently published systematic review found that the majority of drugs for hematologic malignancies are not cost-effective at their current prices (48). Similar trend is observed in other cancer treatments (87, 86, 193). The cost of care has become an important component for delivering high quality-care (163, 20, 21).

We do not recommend that clinicians should choose less effective management strategies; instead we propose that the price of oral-targeted therapies need to be reduced such that the treatment becomes cost-effective and more affordable. Besides price reduction, strategies to optimize drug and dose schedule are needed. Minimal-residual disease (MRD) negative remissions have been reported with drugs such as venetoclax. Clinical trials are needed to ascertain if drug-discontinuation in patients meeting certain parameters (such as MRD negative remission) would be an effective approach. Similar approaches have been utilized in patients with CML who have

received imatinib (STIM trial (147)).

In conclusion, this study provides a comprehensive analysis of the changing prevalence and cost of CLL care in the US Oral targeted therapies will substantially increase survival rates; however, with the current price structure, they will dramatically increase the cost of CLL management for both patients and payers. Such an economic impact could result in financial toxicity, limited access, and lower adherence to the oral therapies, which may undermine their clinical effectiveness. More sustainable pricing strategy is needed for targeted therapies.

CHAPTER III

OPTIMAL LIVER CANCER SURVEILLANCE IN HEPATITIS C-INFECTED POPULATION

3.1 Introduction

Hepatocellular carcinoma (HCC) is the most common type of liver cancer accounting for more than 90% of liver cancer cases (143). The disease burden has been growing at a rapid pace in the past few decades (185). The incidence of HCC in the US has tripled since 1975 (10). Also, while the overall death rate for cancers has declined by about 18% in the past two decades, HCC-related mortality has increased by 40% during the same period (143).

The leading cause of HCC is chronic hepatitis C virus infection, which affects nearly 3 million Americans (61) and contributes to more than 50% of HCC cases in the US (143). Currently, hepatitis C-related HCC is the fastest-growing cause of cancer-related deaths (68). Even though hepatitis C prevalence is declining, HCC mortality is likely to continue to increase for several years (110, 55).

HCC progression is often silent with rare clinical symptoms, and thus HCC is usually fatal or significantly impairs patients' quality of life (207). While treatment is highly effective at earlier stages, no curative options are available at later stages (209, 33). Therefore, early detection is the key to a successful management of HCC patients.

Recent research has shown that regular surveillance with ultrasound in hepatitis C patients can detect HCC at earlier and most curable stages (202). However, unlike many other major cancers such as breast, prostate, and colorectal cancers, surveillance for HCC has been heavily underutilized in practice (204). Currently, only about 10%

of HCC cases are detected at earlier stages and are eligible for curative therapies such as surgical resection, ablation, or liver transplantation (39, 68). As a result, although timely diagnosis and treatment can lead to favorable survival outcomes (5-year survival is about 70% with early detection), most HCC patients have poor prognosis with a 5-year survival of less than 5% (69).

On the other hand, the rapid increase in HCC incidence and mortality, and advancements in detection and treatment in recent years have drawn greater attention and led to wider advocacy for surveillance for HCC (202, 17). However, there is no consensus on the optimal HCC surveillance policies (236, 113). To date, no randomized controlled trial, the clinical gold standard, has assessed the mortality benefits from different HCC surveillance strategies. Further, such trials are considered "unethical" to conduct given that the benefits of ultrasound surveillance have been observed and recognized in clinical practice (174).

In the absence of randomized trials, a comprehensive mathematical can be instrumental in answering several important policy questions. In this study, we propose a general and flexible mathematical modeling framework for chronic disease prevention and treatment problems and utilize it to seek answers to many important health policy questions in the context of HCC surveillance. For example, which hepatitis C patients should be targeted for HCC surveillance? What are the long-term health and economic consequences of HCC surveillance? Should the frequency of surveillance depend on the extent of hepatitis C infection? Should surveillance strategies be adjusted by age? What is the optimal age to terminate surveillance?

In the operations research/management science (OR/MS) literature, chronic disease screening/surveillance problems have been commonly formulated using Markov decision process (MDP) and partially observable MDP (POMDP) modeling frameworks and the optimal policies are computed in an iterative manner based on backward induction-type algorithms (such as value iteration or policy iteration), which we

refer to as the dynamic programming (DP)-based approaches. For example, Chhatwal et al. develop an MDP model to determine the optimal strategy for breast cancer biopsy based on mammogram results and age (46) . Maillart et al. enumerate and evaluates numerous breast cancer screening policies using a partially observable Markov chain (148). Ayer et al. propose a POMDP model to optimize breast cancer screening strategies from individual patients' perspective (18, 19). Zhang et al. determine the optimal biopsy referral decision based on the results of screening tests for prostate cancer using a similar approach (243). Kırkızlar et al. formulate an MDP model to identify the optimal timing of testing and treatment for a class of asymptomatic diseases from a policy cost-effectiveness perspective (119). Yang et al. formulate a dynamic program and approximately solves the optimal age- and gender-specific threshold of biennial screening for childhood obesity to minimize the disease prevalence (237). Rauner et al. consider a bi-objective optimization problem for risk-group based screening for chronic diseases, and utilize a metaheuristic algorithm to determine Pareto-optimal policies (180).

Unless handled specially, solutions to DP-based approaches typically lead to *fully dynamic policies*, policies that have frequent changes of decisions over time. However, in a population-based cancer screening program, frequently changing screening intervals are not practical from a health policy perspective. Instead, policy-makers typically prefer structured policies with less frequently changing intervals. In a DP-based framework, while it is technically possible to impose structured policies with less frequent changes over time by expanding the state space, such an approach usually comes at a cost of significant computational complexity to the extent that it often becomes computationally intractable. Therefore, rather than pursuing such a DP-based formulation and solution route, we instead introduce *M-switch policies* in a mixed integer programming (MIP)-based framework, where the surveillance intervals can change at most M times in a given subpopulation. For example, in a 1-switch

policy (i.e., $M = 1$), surveillance intervals can change at most once for each fibrosis subpopulation (e.g., for patients with advanced fibrosis, screen every 6 months until age 60 and switch to one year afterwards).

Another important feature to be captured as in many disease screening/surveillance problems is that the natural history of the disease is jointly characterized by multiple co-existing conditions, and that the surveillance/screening decisions for one condition can be stratified by the status of other co-existing conditions. For example, in our HCC surveillance problem, disease progression is jointly characterized by observable liver fibrosis states (capturing the stage of hepatitis C infection) and unobservable HCC states (capturing the cancer stage), where liver fibrosis stages affect the cancer risks. In identifying HCC surveillance policies that are stratified by fibrosis stages, we need to ensure that the surveillance policy recommends the same action for patients with the same fibrosis state, irrespective of their unobservable cancer states. For instance, if a policy recommends annual HCC surveillance in advanced fibrosis stage, this policy applies to all patients with advanced fibrosis irrespective of their unknown cancer status (e.g., some may have cancers, and others may not). Mathematically, this implies that while the system is characterized by a tuple of states (e.g., (s_1, s_2)), the optimal action is determined based on only part of them (e.g., optimal actions for (s_1, s_2) and (s_1', s_2') should be the same as long as $s_1 = s_1'$). While such a phenomenon is common in many health policy problems, ensuring it requires expanding the state space and comes at a significantly increased computational cost in a DP-based formulation and solution approach.

In this study, we propose a flexible modeling framework for chronic disease prevention and treatment decisions based on mixed integer programming (MIP), which overcomes the above mentioned limitations of the commonly used DP-based approaches in the chronic disease prevention literature. While in this paper we focus on liver cancer surveillance problem, our modeling framework is general and could be extended

to other population-based disease prevention problems, such as cervical cancer or colorectal cancer screenings, as well as disease screening/surveillance problems when co-morbid conditions exist (e.g., coronary artery disease screening in patients with hypertension and breast cancer screening in pre-diabetic and diabetic patients).

Our contributions are twofold. From the modeling and theoretical perspective, (1) we propose a general modeling framework to ensure practical surveillance recommendations at the policy level, which is applicable to a broad family of disease prevention/treatment problems; and (2) we theoretically analyze the HCC surveillance problem and a) identify the sufficient conditions under which additional surveillance improves the cost-effectiveness of a policy, b) characterize when the surveillance policies should be adapted to populations with different disease progressions, and c) quantify the trade-off between decreasing HCC incidence and increasing treatment outcomes due to advancements in hepatitis C treatment and characterize their net effect on the surveillance policies. From the health policy perspective, we use the best evidence based on large datasets and clinical literature in parameterizing our model and find that (1) unlike the existing guidelines, expanding HCC surveillance to patients in earlier stage of hepatitis C infection improves health outcomes and cost-effectiveness; (2) stratified surveillance policies based on hepatitis C infection stages are more cost-effective than the currently recommended one-size-fits-all type policies; and (3) cost-effectiveness improves when surveillance is adjusted dynamically based on age, and ultimately terminates at a certain age (depending on the hepatitis C infection stages), as continuing surveillance beyond a certain age has minimal outcome benefit.

The remainder of this paper is organized as follows. In Section 3.2, we provide background information on disease dynamics and management. In Section 3.3, we present our model formulation for different classes of surveillance policies, and in Section 3.4, we present our analytical results. In Section 3.5, we describe model

inputs and parameter estimations, and present numerical results. Finally, in Section 3.6, we summarize our findings and conclude.

3.2 Background

Stages of hepatitis C infection (Precursor states of HCC). Chronic hepatitis C infection causes liver inflammation, which leads to a scarring process and forms fibrous scar tissues in liver, called liver *fibrosis*. The extent of fibrosis is graded by the METAVIR scoring system from F0 to F4 stages (23), where F0, F1, and F2 stages represent none, mild, and moderate fibrosis respectively, F3 stage represents advanced fibrosis, and F4 stage represents compensated cirrhosis (CC) (see Table 3.1 for a summary). In CC stage, the entire liver is scarred but still has well-preserved liver functions. Further degradation in liver functionality leads to decompensated cirrhosis (DC), where the disease starts to interfere normal liver functions and may cause many life-threatening complications, including ascites, bleeding varices, and encephalopathy (73). Patients in F3 or more advanced fibrosis stages (i.e., CC and DC) are at risk of developing HCC.

Table 3.1: Overview of stages of hepatitis C infection and liver cancer.

Stages of hepatitis C infection (fibrosis)	
F0	No fibrosis
F1 and F2	Mild and moderate fibrosis, respectively
F3*	Advanced fibrosis without cirrhosis
F4(CC)*	Compensated cirrhosis (CC): the liver is heavily scarred but without serious complications
DC*	Decompensated cirrhosis (DC): the liver is unable to function properly with life-threatening complications
Stages of HCC (BCLC staging system)	
BCLC-0	Very early stage, with single tumor <2cm
BCLC-A	Early stage, single 2-5cm or 3 nodules 3cm, candidate for liver transplant
BCLC-B,C,D	Intermediate, advanced, and terminal stages, respectively, which are not eligible for curative treatments

*: Fibrosis stages that are at risk of developing liver cancer.

Surveillance and diagnosis of HCC. As advocated in practice guidelines, abdominal ultrasonography has been the mainstay test for HCC surveillance (33, 143). While there are other tests such as serological tests like α-fetoprotein (AFP), they are not recommended for surveillance due to inadequate sensitivity (78). Any positive result from an ultrasound surveillance test needs to be further evaluated by diagnostic tests such as computerized tomography (CT), magnetic resonance imaging (MRI), or biopsy. If the diagnostic tests confirm the presence of HCC, the patient receives immediate treatment. Otherwise, the patient undergoes no further intervention until the next surveillance test.

HCC staging and treatments. Staging of HCC is typically determined by the Barcelona-Clinic Liver Cancer (BCLC) staging system (142, 79). Available treatment options include resection, liver transplantation, local ablation (with radiofrequency or percutaneous ethanol injection), and palliative treatment (i.e., chemoembolisation), and selection of the specific treatment depends on HCC and fibrosis stages at diagnosis. For example, patients with advanced stage HCC typically receive palliative treatment such as chemotherapy, patients without decompensation are eligible for liver resection, and liver transplantation is offered only to patients meeting the Milan criteria (a set of criteria characterizing the size of tumors which is used for selecting patients with better prognosis after transplantation) (33, 143).

Currently, annual management cost for hepatitis C-related HCC in the US is estimated to be $1 billion, and is expected to peak to about $1.4 billion in 2025 (181). Given such high economic burden, it is important to provide "care that is based on the wise and cost-effective management of limited clinical resources" (3). As noted in (167), assessing the cost-effectiveness of population-based policies is undoubtedly critical in directing the limited societal health care resources towards the greatest health gains efficiently.

Few simulation-based cost-effectiveness studies in public health literature have

examined HCC surveillance strategies. These studies found that surveillance every 6-12 months is cost-effective in hepatitis C-related cirrhotic patients (137, 51, 13). Based on these pieces of evidence, surveillance every 6-12 months for cirrhotic patients is now recommended in HCC management guidelines, including the guidelines by the American Association for the Study of Liver Disease (AASLD) (33), the European Association for the Study of the Liver (EASL) (143), and the National Comprehensive Cancer Network (NCCN) (24).

However, the existing studies that provided the initial evidence for the current practice guidelines have several common limitations. First, they consider only a very limited number of policies (about 2-3 policies such as no screening, semiannual screening, and annual screening). Second, these studies assume that surveillance should only be offered to patients in an advanced stage of hepatitis C (i.e., cirrhosis), and do not consider the recently emerging evidence from a large clinical trial demonstrating that patients in an earlier stage of hepatitis C (i.e., advanced fibrosis without cirrhosis) may also develop HCC (144). Third, they assume the surveillance interval should be the same in all cirrhotic stages, while the evidence suggests that HCC incidences are different depending on the liver fibrosis stage (144, 73, 173). Lastly, the existing studies only consider a fixed surveillance interval throughout patients' lifetimes (i.e. *routine policies*). However, patients' life expectancy and thus the value of a surveillance strategy is likely to change with age, which suggests that policies adjusted by age may outperform routine policies.

3.3 Model formulation

In this section, we first formulate a stochastic disease progression model to characterize the disease dynamics and clinical process in Section 3.3.1. Then in Section 3.3.2, we provide a base MIP formulation for identifying the optimal surveillance policy. Later, in Section 3.3.3, we extend our base MIP formulation to enforce M-switch

policies, a class of structured policies, that are practical for implementation.

Our target population for HCC surveillance are hepatitis C-infected people who are aware of their infection and fibrosis stage. We take a societal perspective where the objective is to optimize cost-effectiveness of population-based surveillance policies in the target population. The effectiveness, or the health outcome, of a policy is measured in terms of total expected quality-adjusted life years (QALYs), which simultaneously measure quantity and quality of life (85, 64). The cost of a policy includes the costs of surveillance tests, diagnostic tests, fibrosis-related maintenance care, and HCC treatment. We start with providing formal definitions of *cost-effectiveness*, *willingness-to-pay*, and *net monetary benefit*, which are commonly used concepts in the health economics literature.

Definition 3.1 (Drummond et al. 2005) *Let* Π *be the policy space,* $E(\pi)$ *and* $C(\pi)$ *be the total QALYs and costs of a policy* $\pi \in \Pi$*, and* $\pi^0 \in \Pi$ *be the comparator policy.*

- Incremental cost-effectiveness ratio *(ICER) of policy* π *compared with* π^0 *is defined as*

$$\text{ICER} = \Delta C/\Delta E, \quad \textit{where } \Delta C = C(\pi) - C(\pi^0) \textit{ and } \Delta E = E(\pi) - E(\pi^0).$$

- Willingness-to-pay *(WTP) value* $\lambda > 0$ *is defined as the cost that the society is willing to pay for an additional QALY, representing the acceptable cost-effectiveness threshold.*

- *A policy* π *is said to be* more cost-effective *than policy* π^0 *if* $\Delta E > 0,\ \Delta C > 0,$ *and* $\text{ICER} < \lambda$.

- *A policy* π *is said to be* dominated *by policy* π^0 *if* $\Delta E < 0$ *and* $\Delta C > 0$.

- *The net monetary benefit, often referred to as* net benefit *(NB), of policy* π *is defined as* $\text{NB}(\pi) = \lambda E(\pi) - C(\pi)$.

The notion of ICER has been used extensively for comparing two health intervention programs: if a new program can improve health outcomes at a reasonable price which is acceptable to the society (i.e., maintaining ICER below the WTP threshold), it is considered more cost-effective than the comparator. NB is later proposed by Stinnett and Mullahy (208). The policy with the maximum NB is known to be the most cost-effective one ((74), also see a formal proof as Lemma B.1 in Appendix B.3.2), and therefore NB has been commonly used to choose the most cost-effective intervention among multiple alternatives.

3.3.1 Disease natural history model

In line with the published studies in clinical literature (51, 13), we model the disease progression as a discrete-time finite-horizon Markov process. We assume that surveillance frequency is at most every 3-month, the minimum surveillance interval considered in clinical practice (236). Let $k \in \mathscr{N} = \{1, 2, \cdots, N\}$ be the time index where each period represents three months. The components of the Markov model are defined as follows.

- $f \in \mathscr{F} = \{0, 1, 2\}$: Fibrosis state, which reflects the severity of degradation in liver functioning and can be easily assessed by non-invasive tests (40). Specifically, 0 represents advanced fibrosis (F3), 1 represents CC, and 2 represents DC. Hereinafter, we refer to the patients with fibrosis state f as the *f-subpopulation*.

- $h \in \mathscr{H} = \{0, 1, 2, 3\}$: Tumor state, representing HCC (i.e., tumor, or cancer) stage, where 0 represents cancer-free state, and 1, 2, and 3 represent small, medium, and large HCC, respectively. As noted earlier, HCC state definitions are based on the BCLC staging system: small HCC ($h = 1$) corresponds to very early stage (BCLC-0), medium HCC ($h = 2$) corresponds to early stage (BCLC-A), and large HCC ($h = 3$) corresponds to late stage (BCLC-B/C/D).

- $\mathscr{S} = \mathscr{S}_d \cup \mathscr{S}_e$: State space, including disease states $\mathscr{S}_d$ and absorbing states $\mathscr{S}_e$. Disease states, $\mathscr{S}_d = \{(f,h) : f \in \mathscr{F}, h \in \mathscr{H}\}$, consist of fibrosis state (f) and tumor state (h). Absorbing states, $\mathscr{S}_e = \{\mathsf{PTX}, \mathsf{DLD}, \mathsf{DOC}\}$, include post-treatment state (PTX), death from liver disease (DLD), and death from other causes (DOC).

- $\mathsf{P}^{|\mathscr{S}|}$: State distribution space, which is a probability simplex of $|\mathscr{S}|$ dimensions and includes all possible distributions over the state space $\mathscr{S}$.

- $a \in \mathscr{A}_s$: Available action to take in state $s \in \mathscr{S}$. We let $\mathscr{A}_s = \{\mathsf{W}, \mathsf{E}\}$ for $s \in \mathscr{S}_d$, where W and E represent *Wait* and *Examine (Screen)* actions, respectively, and $\mathscr{A}_s = \{\mathsf{W}\}$ for $s \in \mathscr{S}_e$, as there is no decision in absorbing states.

- π: HCC surveillance policy, which specifies screening actions to be taken for each f-subpopulation at every time period, given the underlying state distribution in the population, i.e., $\pi : \mathsf{P}^{|\mathscr{S}|} \times \mathscr{N} \mapsto \{\mathsf{W}, \mathsf{E}\}^{|\mathscr{F}|}$. For example, given the distribution of the tumor states, a surveillance policy may recommend no screening in F3 and CC subpopulations and screening in DC subpopulation in a given period. In order to compute the HCC distribution at any given time period, it is sufficient to know the initial state distribution and prior sequence of actions, because the distribution in each subsequent period only depends on the action taken and the distribution at the previous period. As such, policy π is reduced to a function that maps the triplet of an initial HCC distribution, a time period, and an f-subpopulation to an action, i.e., $\pi : \mathsf{P}^{|\mathscr{S}|} \times \mathscr{N} \times \mathscr{F} \mapsto \{\mathsf{W}, \mathsf{E}\}$. Properties and structures of surveillance policies are discussed in more detail in Section 3.3.3.

- $O(o|(f,h)), o \in \{+,-\}$: Observation probability, representing the probability of observing test outcome o given the true underlying disease state (f,h). The test outcome o represents the final result of a two-step surveillance-diagnostic

process, where o is positive (+) if both surveillance and diagnostic tests are positive. Observation probabilities are specified by test accuracies, which are characterized by sensitivity (the probability of a positive outcome given that the patient has HCC) and specificity (the probability of a negative outcome given that the patient is cancer-free). Let $sens_s$ and $sens_d$ denote the sensitivity of the surveillance and diagnostic tests, respectively. Similarly, let $spec_s$ and $spec_d$ denote the specificity of surveillance and diagnostic tests, respectively. We assume perfect specificity for diagnostic tests, which is reasonable because specificity of HCC diagnosis tests is typically greater than 95% (188, 133) and a false positive diagnosis of liver cancer is rare in practice. Then, for any $f \in \mathscr{F}$, we can compute observation probabilities as follows:

$$O\left(o|(f,h)\right) = \begin{cases} sens_s sens_d, & \text{if } o = +,\ h \in \{1,2,3\} \\ 1 - sens_s sens_d, & \text{if } o = -,\ h \in \{1,2,3\} \\ (1 - spec_s)(1 - spec_d) = 0, & \text{if } o = +,\ h = 0 \\ spec_s + (1 - spec_s) spec_d = 1, & \text{if } o = -,\ h = 0 \end{cases}$$

- $P_k(s'|(f,h),\mathsf{W}), \forall (f,h) \in \mathscr{S}_d, s' \in \mathscr{S}$: Natural progression probability, i.e., the probability of transition from state $(f,h) \in \mathcal{S}_d$ in period k to $s' \in \mathscr{S}$ in the next period, given the action $a = \mathsf{W}$. For transitions to death states, $P_k(\mathsf{DLD}|(f,h),\mathsf{W})$ represents excess mortality due to specific liver disease state (f,h) and $P_k(\mathsf{DOC}|(f,h),\mathsf{W})$ represents background mortality (i.e., other-cause mortality).

- $P_k(s|s,\mathsf{W}) = 1, \forall s \in \mathscr{S}_e$: Absorbing states have self-transition probability of 1.

- $P_k(s'|(f,h),\mathsf{E})$: Transition probability that a patient will be in state $s' \in \mathscr{S}$ in period $k+1$, given that he is in state $(f,h) \in \mathscr{S}_d$ and takes *Screen* action in period k. Cancer-free patients (i.e., $h = 0$) do not need treatment, and thus the

state transitions simply follow the natural progression, i.e., $P_k(s'|(f,0),\mathsf{E}) = P_k(s'|(f,0),\mathsf{W}),\ \forall f \in \mathscr{F},\ s' \in \mathscr{S}$. For patients with HCC (i.e., $h \in \{1,2,3\}$), either the tumor could be detected and patients then receive immediate treatment, i.e., $P_k(\mathsf{PTX}|(f,h),\mathsf{E}) = O(+|(f,h)) = sens_s sens_d$; or the tumor could be missed by screening tests and hence patients follow the natural history until the next surveillance test, i.e., $P_k(s'|(f,h),\mathsf{E}) = O(-|(f,h))P_k(s'|(f,h),\mathsf{W}) = (1 - sens_s sens_d)P_k(s'|(f,h),\mathsf{W}),\ \forall f \in \mathscr{F},\ s' \neq \mathsf{PTX}$ (see Figure 3.1 for an illustration of state transitions).

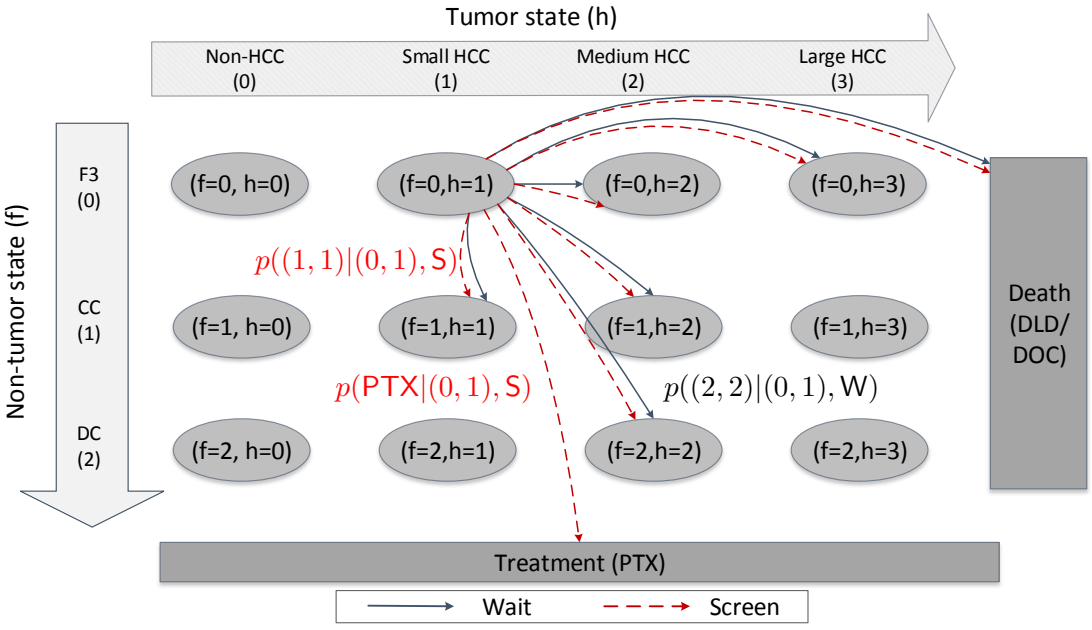


Figure 3.1: State transition diagram.

Note. Health states are two dimensional, including fibrosis state and tumor state; absorbing states include death states and post-treatment (PTX) state. Progression can occur in fibrosis and tumor states simultaneously in one period. Transition to post-treatment state is possible only when the action is screening E.

- $q_k(s,a,s')$ and $c_k(s,a,s'),\ \ k = 1,\cdots,N-1$: Intermediate QALYs and costs, respectively, for taking action a, given that a patient is in state s at period k and transits to state s' in period $k+1$. A *Screen* action incurs additional costs of surveillance and possible diagnostic tests. In particular, with respective

probabilities $(1 - spec_s)$ and $sens_s$, a cancer-free patient and a cancer patient will undergo diagnostic tests to confirm the surveillance test results. Therefore, we define the test costs for cancer-free patients (i.e., $h = 0$) as $\kappa_0 := c_s + (1 - spec_s)c_d$, and for patients with missed cancers (i.e., $h \geq 1$ and $s' \neq \mathsf{PTX}$) as $\kappa_1 := c_s + sens_s c_d$, where c_s and c_d represent the cost of surveillance and diagnostic tests, respectively. We remark that while specificity of surveillance test does not affect state transitions, it affects cost calculations because lower specificity (i.e., higher false positive rate) leads to more costly further diagnostic tests. Thus, when either the patient is cancer-free or his cancer is missed by screening and diagnostic tests (i.e., $s' \neq \mathsf{PTX}$), the expected intermediate costs for taking action *Screen* is equal to:

$$c_k((f,h),\mathsf{E},s') = \begin{cases} c_k((f,h),\mathsf{W},s') + \kappa_0, & \text{if } h = 0, \\ c_k((f,h),\mathsf{W},s') + \kappa_1, & \text{if } h \geq 1. \end{cases}$$

When $s' = \mathsf{PTX}$, $q_k(s,\mathsf{E},\mathsf{PTX})$ and $c_k(s,\mathsf{E},\mathsf{PTX})$ represent the lump-sum QALYs from treatment capturing the cumulative remaining life expectancy adjusted by quality of life under treatment, and the cumulative cost of surveillance, diagnosis, and treatment, respectively.

- $q_k(s,a)$ and $c_k(s,a)$, $k = 1, \cdots, N-1$: Expected QALYs and costs accumulated in state s when action a is taken in period k, which are respectively computed as $q_k(s,a) = \sum_{j \in \mathscr{S}} q_k(s,a,j) P_k(j|s,a)$ and $c_k(s,a) = \sum_{j \in \mathscr{S}} c_k(s,a,j) P_k(j|s,a)$, for $k = 1, \cdots, N-1$.

- $q_N(s)$ and $c_N(s)$: Terminal QALYs and costs in state $s \in \mathscr{S}$, respectively.

- $r_k(s,a)$ for $k = 1, \cdots, N-1$ and $r_N(s)$: Expected NB accumulated in period k given state s and action a, and the terminal NB for state s, respectively. The intermediate NB is computed as $r_k(s,a) = \lambda q_k(s,a) - c_k(s,a)$ for $k =$

$1, \cdots, N-1$, and the terminal NB is computed as $r_N(s) = \lambda q_N(s) - c_N(s)$.

3.3.2 The base model

In a proper HCC surveillance policy, patients in a given f-subpopulation should be recommended the same screening action. That is, patients with the same fibrosis state f should share the same screening action, regardless of their true tumor states. This implies that the surveillance policy inherently depends on the underlying distribution of HCC states (i.e., the cancer prevalence). Thus, our problem in principle can be formulated as a continuous-state dynamic program, and be considered a special case of a POMDP without belief updates. However, POMDP models for problems of our size suffer from the "curse of dimensionality" and are notoriously difficult to solve. Indeed, in Appendix B.1, we develop and present such a POMDP model formulation for our problem and report its performance. As we show there, suffering from the curse of dimensionality, even a simplified POMDP model without the M-switch policy structure cannot be solved for more than a few iterations.

Hence, we turn our attention to alternative modeling approaches and instead propose an MIP-based model formulation, which naturally allows formulation of structured policies without losing computational tractability. It is well-known that sequential decision making problems can be modeled and solved using linear programming (LP) (112, 11). Building upon this LP-based modeling framework, we first present a base HCC surveillance (HS) MIP model , which ensures that actions are adapted to f-subpopulations, irrespective of unobservable tumor states, as follows. Later in Section 3.3.3, we introduce the construction of M-switch policy structures and the corresponding model formulation.

$$\text{(HS)} \quad \max_{(x_k,y_k)_{1\le k\le N}} \quad NB(\mathbf{x}) = \sum_{k=1}^{N-1}\sum_{s\in\mathscr{S}}\sum_{a\in\mathscr{A}_s} r_k(s,a)x_k(s,a) + \sum_{s\in\mathscr{S}} r_N(s)x_N(s) \tag{3.1}$$

$$s.t. \quad \gamma\sum_j\sum_{a\in\mathscr{A}_j} P_k(s|j,a)x_k(j,a) = \sum_{a\in\mathscr{A}_s} x_{k+1}(s,a),$$

$$\forall s\in\mathscr{S}, k=1,\cdots,N-1, \tag{3.2}$$

$$\gamma\sum_j\sum_{a\in\mathscr{A}_j} P_{N-1}(s|j,a)x_{N-1}(j,a) = x_N(s), \quad \forall s\in\mathscr{S}, \tag{3.3}$$

$$\sum_{a\in\mathscr{A}_s} x_1(s,a) = \alpha(s), \quad \forall s\in\mathscr{S} \tag{3.4}$$

$$x_k(s,a) \le y_k(s,a), \quad \forall s\in\mathscr{S}, a\in\mathscr{A}_s, k=1,\cdots,N-1, \tag{3.5}$$

$$\sum_{a\in\mathscr{A}_s} y_k(s,a) = 1, \quad \forall s\in\mathscr{S}, k=1,\cdots,N-1, \tag{3.6}$$

$$y_k((f,h),a) = y_k((f,h'),a),$$

$$\forall f\in\mathscr{F}, h,h'\in\mathscr{H}, \forall a\in\mathscr{A}_{(f,h)}, k=1,\cdots,N-1, \tag{3.7}$$

$$x_k(s,a), x_N(s) \ge 0, \quad \forall s\in\mathscr{S}, a\in\mathscr{A}_s, k=1,\cdots,N-1, \tag{3.8}$$

$$y_k(s,a)\in\{0,1\}, \quad \forall s\in\mathscr{S}, a\in\mathscr{A}_s, k=1,\cdots,N-1, \tag{3.9}$$

where $\alpha(s)$ is the initial distribution of state s, and γ is the discount factor. $x_k(s,a)$ is the occupancy measure (11), which can be interpreted as the (discounted) joint probability of being in state s and choosing action a in period k, i.e., if we let random variables S_k and A_k to denote the state and action at time t, then $x_k(s,a) = \gamma^{k-1}\mathbb{P}(S_k = s, A_k = a)$. The objective function (3.1) maximizes the expected total discounted NB. Constraints (3.2) and (3.3) represent the flow balance equations capturing the dynamic updates for the occupancy measures over time, which can be easily verified by observing that $\gamma\sum_j\sum_{a\in\mathscr{A}_j} P_k(s|j,a)x_k(j,a) = \gamma\sum_j\sum_{a\in\mathscr{A}_j}\mathbb{P}(S_{k+1}=s|S_k = j, A_k = a) = \gamma^k\mathbb{P}(S_{k+1}=s) = \sum_{a\in\mathscr{A}_s}\gamma^k\mathbb{P}(S_{k+1}=s, A_{k+1}=a) = \sum_{a\in\mathscr{A}_s} x_{k+1}(s,a)$. Constraint (3.4) captures the initial condition for the system dynamics.

We consider only deterministic (i.e., non-randomized) policies because randomized

policies (e.g., screen 50% of the CC patients) are not practical to implement. Let $y_k(s, a)$ be an indicator variable, which equals to 1 if action a is chosen at state $s \in \mathscr{S}$ in period k, and 0 otherwise. Then, Constraint (3.5) links the occupancy measures and indicator variables, and Constraint (3.6) ensures the policy to be deterministic. Lastly, Constraint (3.7), which we call the *state-linking equation*, ensures that the same action is taken in the same fibrosis state, irrespective of the tumor states.

3.3.3 Construction of M-switch policies and the full model

The solution to HS leads to a *fully dynamic policy*, where intervals between surveillance tests may be irregular and may change frequently over time. For example, in a fully dynamic policy as illustrated in Figure 3.2, a specific f-subpopulation may be recommended to undergo surveillance following 6, 3, 12, and then again 6 months surveillance intervals subsequently.

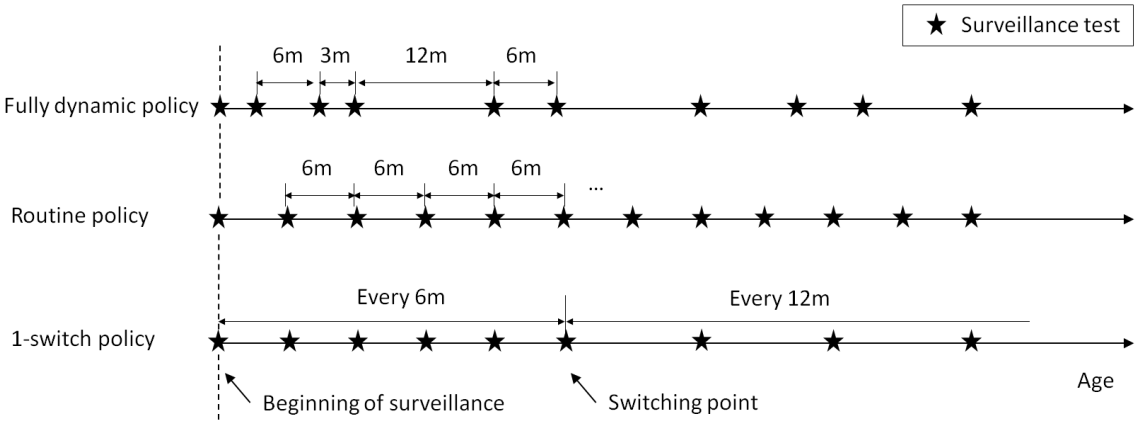


Figure 3.2: A schematic diagram of different types of policies for one f-subpopulation.

Note. 6m represents 6-month, and so forth.

While from a theoretical point of view the fully dynamic policy is optimal among all feasible HCC surveillance policies, such a policy would be difficult to implement due to frequently changing surveillance intervals. In contrast, *routine policies* simply recommend a single fixed surveillance interval for all patients throughout their lifetimes, irrespective of their fibrosis stage and age (see Figure 3.2). While routine

policies are practical due to their simplistic form, they, on the other hand, may not well capture the dynamics of disease progression over time, and thus may have a poor performance.

To balance the effectiveness of the fully dynamic policy and the practicality of routine policies, we consider another class of structured policies, which we call *M-switch policies*. In M-switch policies, we allow switches in surveillance intervals but explicitly limit the total number of switches. Specifically, for each f-subpopulation, surveillance intervals are allowed to change at most M times until surveillance terminates in that f-subpopulation (see Figure 3.2). For instance, in a 1-switch policy, surveillance intervals for CC subpopulation can be 6 months until age 60, and switch to one year afterwards. We remark that while we optimally determine the terminal age, terminating surveillance is not counted as an additional switch in an M-switch policy.

We are especially interested in two special cases where $M = 0$ or $M = 1$. In a 0-switch policy (i.e., $M = 0$), surveillance intervals do not change until the surveillance terminates, but they are stratified by fibrosis states and can be different across f-subpopulations. Therefore, we call such policies *stratified policies*. A 1-switch policy (i.e., $M = 1$) can have at most one switch in surveillance intervals until surveillance terminates for each f-subpopulation, which enables the surveillance strategy to consider and adapt to the aging effects in the population. Yet, compared with a fully dynamic policy, such a policy has much fewer changes in surveillance intervals, and hence is easier to follow. Maillart et al. call 1-switch policies as "two-phase policies" and compares more than 1200 such policies via explicit enumeration (148).

Remark 3.1 *Although M-switch policies have a much simpler structure compared with the fully dynamic policies, the total number of M-switch policies is still prohibitively large for explicit enumeration within a reasonable computation time. For example, to determine a 1-switch policy in a practical setting which we shall specify*

shortly, we will choose the switching and terminating period from $\frac{100-40}{4}$ *possible time points (i.e., every 4 years from age 44 to 100), select 2 different surveillance intervals from a list of 5 intervals for the periods before and after the switching period respectively, and repeat such selection for all 3 different* f*-subpopulations. Thus, the total number of feasible 1-switch policies is* $\left(\binom{(100-40)/4}{2} \times 5 \times 4\right)^3 \sim 9.3 \times 10^9$*. Assuming that evaluating each policy takes about 0.005 seconds, it takes* $9.3 \times 10^9 \times 0.005/(3600 \times 24)$ *days* > 1.4 *years to evaluate all such possible policies.*

Now we introduce the mathematical formulation for M-switch policies. Let d be the screening interval, and $\mathscr{D}$ be the set of possible screening intervals. For no surveillance, we define screening interval $d = \infty$. Surveillance intervals can only change every C periods, called a *switching cycle*, which is a common multiple of all possible intervals $\{d \in \mathscr{D} : d \neq \infty\}$. Let $l \in \mathscr{L} = \{1, 2, \cdots, \lceil \frac{N-1}{C} \rceil\}$ be the index for the switching cycles and $\mathscr{J}_i$ denote the set of screening periods with interval length $d_i \in \mathscr{D}$. Then,

$$\mathscr{J}_i = \begin{cases} \{1 + nd_i : n = 0, 1, \cdots, \frac{C}{d_i} - 1\} & \text{for } d_i < \infty, \\ \emptyset & \text{for } d_i = \infty. \end{cases}$$

3.3.3.1 Formulation with nested structure of screening intervals

Because practical cancer screening strategies typically consider every 3-month, semiannual, annual, and biennial screenings, they naturally form a "nested structure". That is, screening periods $\mathscr{J}_1 \supset \mathscr{J}_2$ if screening interval $d_1 \leq d_2$. For example, Figure 3.3 illustrates the screening periods $\mathscr{J}$ for $d \in \mathscr{D} = \{d_1 = 1, d_2 = 2, d_3 = 4, d_4 = 8, d_5 = 16, d_6 = \infty\}$, which represent three-month, semiannual, annual, biennial, four-year, and no surveillance, respectively, with a cycle length $C = 16$ (i.e., 4 years). In this section, we first present the model formulation for cases where such a nested structure exists. Later, in Section 3.3.3.2, we extend this and present a more general model formulation.

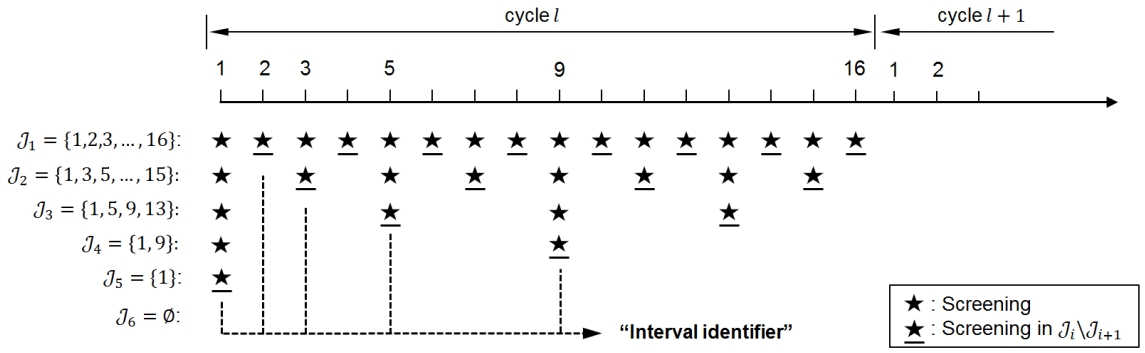


Figure 3.3: Surveillance periods for different surveillance intervals in one switching cycle.

Clearly, we observe $\mathscr{J}_1 \supset \mathscr{J}_2 \supset \cdots \supset \mathscr{J}_6$, which implies that surveillance periods in a less aggressive policy (i.e., a policy with larger surveillance intervals) is a subset of those in a more aggressive policy. This nested structure implies that given a policy $\mathscr{J}_{i+1}$, we can construct a more aggressive policy $\mathscr{J}_i$ by adding all surveillance periods in $\mathscr{J}_i \backslash \mathscr{J}_{i+1}$ to $\mathscr{J}_{i+1}$ (see Figure 3.3, where the set of surveillance period $\underline{\bigstar}$'s in each row i corresponds to $\mathscr{J}_i \backslash \mathscr{J}_{i+1}$). For example, we can construct semiannual surveillance policy $\mathscr{J}_2$ by adding additional surveillance at periods $\mathscr{J}_2 \backslash \mathscr{J}_3$ to annual surveillance policy $\mathscr{J}_3$. We can generalize this as $\mathscr{J}_i = \cup_{i'=i}^{5} (\mathscr{J}_{i'} \backslash \mathscr{J}_{i'+1})$ for $i = 1, \cdots, 5$, where $(\mathscr{J}_{i'} \backslash \mathscr{J}_{i'+1})$'s are disjoint.

We let $\mathscr{I}$ denote the *set of interval identifiers*, which means that the surveillance interval d in one cycle is uniquely determined by the actions at periods in $\mathscr{I}$. In particular, given $\{\mathscr{J}_i\}$ considered in this problem, we have $\mathscr{I} = \{1, 2, 3, 5, 9\}$ (see Figure 3.3). For example, for any given cycle $l \in \mathscr{L}$, the case that $y_{(l-1)C+j}((f,h), \mathsf{E}) = 1$ for $j \in \{1, 5, 9\}$ and $y_{(l-1)C+j'}((f,h), \mathsf{E}) = 0$ for $j' \in \{2, 3\}$ represents every 12-month surveillance in f-subpopulation.

Based on these arguments, we first introduce the following two constraints, which

ensure that each switching cycle is designated to a specific surveillance interval.

$$y_{(l-1)C+j}((f,h),\mathsf{E}) \le y_{(l-1)C+j'}((f,h),\mathsf{E}),$$
$$\forall l \in \mathscr{L},\ f \in \mathscr{F}, h \in \mathscr{H},\ j \in \mathscr{J}_i \backslash \mathscr{J}_{i+1}, j' \in \mathscr{J}_{i+1}, i = 1,2,3,4, \quad (3.10)$$
$$y_{(l-1)C+j}((f,h),\mathsf{E}) = y_{(l-1)C+j'}((f,h),\mathsf{E}),$$
$$\forall l \in \mathscr{L},\ f \in \mathscr{F}, h \in \mathscr{H},\ j, j' \in \mathscr{J}_i \backslash \mathscr{J}_{i+1}, i = 1,2,3,4,5, \quad (3.11)$$

where Constraint (3.10) reflects the nested structure of $\{\mathscr{J}_i\}$: if we choose any surveillance period in $\mathscr{J}_i \backslash \mathscr{J}_{i+1}$, we must also choose all the periods in $\mathcal{J}_{i+1}$. Constraint (3.11) ensures that the same action is taken in all periods in $\mathscr{J}_i \backslash \mathscr{J}_{i+1}$. We remark that if $\frac{N-1}{C}$ is not integral, Constraints (3.11) and (3.10) hold for periods $k \le N-1$ in the last (incomplete) cycle $l = \lceil \frac{N-1}{C} \rceil$.

Next, we formulate the number of interval switches using additional constraints. Since terminating surveillance (i.e., switching to no screening $\mathscr{J}_6 = \emptyset$ in the next cycle) is not counted as a separate switch, we first need to determine whether surveillance terminates in the next period or not. For this purpose, we define a binary variable $z_{l,f}$ and let it equal 0 if f-subpopulation has no surveillance in cycle l, and 1 otherwise. We remark by Constraint (3.10), $y_{(l-1)C+1}((f,h),\mathsf{E}) = 0$ implies that actions in all periods in cycle l are *Wait*. Therefore, to determine the value of $z_{l,f}$, it is sufficient to check the value of $y_{(l-1)C+1}((f,h),\mathsf{E})$. That is,

$$z_{l,f} = y_{(l-1)C+1}((f,h),\mathsf{E}), \quad \forall l = 1,\cdots,|\mathscr{L}|, f \in \mathscr{F}, h \in \mathscr{H}, \quad (3.12)$$
$$z_{l,f} \in \{0,1\}, \quad \forall l = 1,\cdots,|\mathscr{L}|, f \in \mathscr{F}. \quad (3.13)$$

Also, to determine the change of surveillance intervals across cycles, we define a binary variable $w_{l,f}$ and let it equal 1 if surveillance interval for the f-subpopulation in cycle l switches to a different *non-stopping* interval in the next cycle, and 0 otherwise.

Then,

$$w_{l,f} - (1 - z_{l+1,f}) \leq \sum_{j \in \mathscr{I}} |y_{(l-1)C+j}((f,h),\mathsf{E}) - y_{lC+j}((f,h),\mathsf{E})|$$

$$\leq |\mathscr{I}|(w_{l,f} + 1 - z_{l+1,f}),\ \forall l = 1, \cdots, |\mathscr{L}| - 1, f \in \mathscr{F}, h \in \mathscr{H}, \quad (3.14)$$

$$\sum_{l=1}^{|\mathscr{L}|-1} w_{l,f} \leq M, \quad \forall f \in \mathscr{F}, \quad (3.15)$$

$$y_{(l-1)C+1}((f,h),\mathsf{E}) \geq y_{lC+1}((f,h),\mathsf{E}), \quad \forall l = 1, \cdots, |\mathscr{L}| - 1, f \in \mathscr{F}, h \in \mathscr{H}, \quad (3.16)$$

$$w_{l,f} \in \{0,1\}, \quad \forall l = 1, \cdots, |\mathscr{L}| - 1, f \in \mathscr{F}. \quad (3.17)$$

In Constraint (3.14), the set $\mathscr{I}$ represents the set of interval identifiers that we defined earlier. Constraint (3.14) becomes redundant if surveillance stops in cycle $l+1$ (i.e., $z_{l+1,f} = 0$); otherwise, it ensures that $w_{l,f} = 1$ if and only if the screening actions in the interval identifier periods of cycle l do not exactly match with those in the next cycle $l+1$ (i.e., surveillance interval in cycle l changes to a different *non-stopping* surveillance interval in cycle $l+1$). Constraint (3.15) explicitly restricts that intervals can switch at most M times for each fibrosis state f. Lastly, Constraint (3.16) ensures that once surveillance terminates at some period, there will be no more surveillance thereafter. Note that Constraint (3.14) can be linearized by change of variables using $x = x^+ - x^-$ and $|x| = x^+ + x^-$ where $x^+, x^- \geq 0$.

Now we can find the optimal M-switch policy (with nested intervals) by solving the following formulation:

$$\text{(HS-M-NI)} \quad \max_{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{w}} \Big\{ NB(\mathbf{x}) \quad s.t.\ (3.2) - (3.17) \Big\} \quad (3.18)$$

3.3.3.2 General formulation without nested structure of screening intervals

While the screening intervals in our problem have a nested structure, in the following, we provide a more generalized formulation, which do not require such a nested structure, and can be useful in some other problems.

As stated earlier, the cycle length C needs to be a common multiple of all possible intervals $\{d \in \mathscr{D} : d \neq \infty\}$. Since the nested structure is not assumed here, it is not sufficient to have $C = \max\{d \in \mathscr{D} : d \neq \infty\}$. For example, the minimal C is 4 years if possible intervals include 0.5, 1, 2, and 4 years; whereas it is 6 years if possible intervals include 0.5, 1, 2, and 3 years.

We define binary variable $\delta^f_{l,i}$ to be 1 if screening interval is $d_i \in \mathscr{D}$ for cycle $l \in \mathscr{L}$ in f-subpopulation, and 0 otherwise. We have the following constraints for screening interval assignment in each cycle l:

$$y_{(l-1)C+j}((f,h), a = \mathsf{E}) \geq \delta^f_{l,i},$$
$$\forall f \in \mathscr{F},\ h \in \mathscr{H},\ j \in \mathscr{J}_i,\ i \in \{1, \cdots, |\mathscr{D}|\}, \tag{3.19}$$
$$y_{(l-1)C+j}((f,h), a = \mathsf{E}) \leq 1 - \delta^f_{l,i},$$
$$\forall f \in \mathscr{F},\ h \in \mathscr{H},\ j \notin \mathscr{J}_i,\ i \in \{1, \cdots, |\mathscr{D}|\}, \tag{3.20}$$
$$\sum_{i=1}^{|\mathscr{D}|} \delta^f_{l,i} = 1, \quad \forall\ f \in \mathscr{F},\ l \in \mathscr{L}, \tag{3.21}$$
$$\delta^f_{l,i} \in \{0,1\}, \quad \forall\ f \in \mathscr{F},\ l \in \mathscr{L},\ i \in \{1, \cdots, |\mathscr{D}|\}. \tag{3.22}$$

Constraints (3.19) and (3.20) ensure that, if interval d_i is selected for f-subpopulation in cycle l (i.e., $\delta^f_{l,i} = 1$), periods of $\mathscr{J}_i$ in that cycle must have screening action (E), and the other periods in that cycle have no screening (W). Both constraints will become redundant if $\delta^f_{l,i} = 0$. Constraint (3.21) guarantees that one and only one interval is selected for cycle l.

To determine whether the screening interval switches from cycle l to $l+1$, we only need to check the values of $\delta^f_{l,i}$ and $\delta^f_{l+1,i}$, and have the following constraints:

$$w_{l,f} - (1 - z_{l,f}) \leq \frac{1}{2} \sum_{i=1}^{|\mathscr{D}|} |\delta^f_{l,i} - \delta^f_{l+1,i}| \leq w_{l,f} + (1 - z_{l,f}),\ \forall l = 1, \cdots, |\mathscr{L}| - 1, f \in \mathscr{F}, h \in \mathscr{H}. \tag{3.23}$$

Observe that $\frac{1}{2} \sum_{i=1}^{|\mathscr{D}|} |\delta^f_{l,i} - \delta^f_{l+1,i}|$ only takes two values: it equals 1 if the screening interval switches from cycle l to $l+1$, and 0 otherwise. When $z_{l,f} = 1$ (i.e., surveillance

will not stop in the next cycle), Constraint (3.23) is reduced to $w_{l,f} = \frac{1}{2}\sum_{i=1}^{|\mathscr{D}|} |\delta_{l,i}^{f} - \delta_{l+1,i}^{f}|$ which is trivially true by the definition of $w_{l,f}$; when $z_{l,f} = 0$, Constraint (3.23) becomes redundant as stopping surveillance (i.e., switching to $d = \infty$) is not counted as a switch.

In addition to these constraints, as in the nested formulation, we also need Constraints (3.12)-(3.13) to determine whether surveillance is "empty" in the next period $l+1$ for fibrosis state f, and Constraints (3.15)-(3.17) to bound the number of switches.

To sum up, the general formulation for M-switch policy without nested structure of screening intervals becomes as follows:

$$\text{(HS-M-G)} \quad \max_{\mathbf{x},\mathbf{y},\delta,\mathbf{w},\mathbf{z}} \Big\{ NB(\mathbf{x}) \quad s.t.\ (3.2)-(3.9), (3.12)-(3.13), \\ (3.15)-(3.17), (3.19)-(3.23) \Big\} \tag{3.24}$$

In Appendix B.2, we show that this general formulation can be reduced to the nested-structure formulation presented in (3.18), when the nested structure in screening intervals exist.

Remark 3.2 *Using our modeling framework, we can also formulate another class of practical surveillance policies,* interval-switching policies*, by choosing a long switching cycle (e.g., 10 years or 20 years) and not specifying the limits on the number of switches (i.e., by removing Constraint* (3.12)-(3.17)*). This class of structured policies are suitable when policy-makers are interested in identifying age group-based surveillance policies, such as determining the surveillance policies for patients of age 40-49, 50-59, and each subsequent 10-year age groups.*

3.4 Analytical Results

In this section, we explore structural properties of the HCC surveillance problem, which may provide useful insights for decision-makers in developing effective surveillance policies. We have three main results: 1) we identify the sufficient conditions

under which additional surveillance improves the cost-effectiveness of a policy (Proposition 3.1); 2) we characterize when the surveillance policies should be adapted to populations with different disease progressions (e.g., genotype 3 hepatitis C virus infection has faster fibrosis progression than other genotypes, see Theorem 3.1); and 3) we quantify the trade-off between decreasing HCC incidence and increasing treatment outcomes (as is happening in current practice), and characterize their net effect on the surveillance policies (Theorem 3.2). All proofs are provided in Appendix B.3.

We start with introducing additional notations which are used in our analytical results. We first define augmented tumor states $\overline{\mathscr{H}} = \mathscr{H} \cup \{4\}$ where state 4 represents all absorbing states as one aggregated state. Hence, the state space defined in Section 3.3 is equivalent to $\mathscr{S} = \mathscr{F} \times \overline{\mathscr{H}}$. Throughout this section, we refer to h as an augmented tumor state $h \in \overline{\mathscr{H}}$ unless its range is explicitly specified. As before, subscript k represents the time index in all of the following notations.

$\overline{\mathscr{H}} = \mathscr{H} \cup \{4\}$	Augmented tumor states where state 4 represents all absorbing states as one aggregated state
$\mathscr{H}^+ = \{1, 2, 3\}$	Cancer states, including small, medium, and large HCC tumors
$\mathbf{F}_k = [F_k(f'\|f)]$	Transition matrix for progression in fibrosis states
$\mathbf{H}_k(f) = [H_k(h'\|h, f)]$	Transition matrix for progression in augmented tumor states $h \in \overline{\mathscr{H}}$ given the fibrosis state f
$R_k(f, h)$	Lump-sum NB of HCC treatment for state (f, h), $h \in \mathscr{H}^+$
$W_k(f, h)$	Lump-sum NB for state (f, h), for $h \in \mathscr{H}^+$ if treatment is delayed to the next period
$\Delta Q_k(f, h)$, $\Delta C_k(f, h)$	QALY and cost components in the incremental net benefit of immediate treatment, respectively, i.e., $R_k(f, h) - W_k(f, h) = \lambda \Delta Q_k(f, h) - \Delta C_k(f, h)$
$V_k^\pi(f, h)$	Value function of state (f, h) for given policy π
$b_k(f, h)$	State distribution over $\mathscr{F} \times \overline{\mathscr{H}}$ (i.e., percentage of population in state (f, h))
$\zeta = sens_s sens_d$	Probability that a cancer will be detected and diagnosed
$\Delta_t(f)$	Marginal surveillance benefit for f-subpopulation at period t

Several useful properties for some of the above quantities are presented in the preliminaries in Appendix B.3.1, which are used in proving some key results in this section.

Throughout this section, we make the following reasonable assumptions, which are in line with the evidence from the medical literature and our data analysis.

Assumption 3.1 *The joint transition probability for natural progression* $P_k((f',h')|(f,h),\mathsf{W})$ *is specified by* $P_k((f',h')|(f,h),\mathsf{W}) = H_k(h'|h,f)F_k(f'|f)$, $\forall f, f' \in \mathscr{F}$, $h, h' \in \overline{\mathscr{H}}$.

This assumption implies that tumor progression depends on fibrosis stage, but not vice versa (144, 173).

Assumption 3.2 *For each disease state* $(f,h) \in \mathscr{S}_d$ *and* $k = 1, \cdots, N-1$, *treatment NB* $R_k(f,h)$, *intermediate NB* $r_k((f,h),\mathsf{W})$, *and terminal NBs* $r_N(f,h)$ *and* $R_N(f,h)$ *have the following properties:*

1. *Non-negativity,*
2. *Monotonicity: decreasing in* $f \in \mathscr{F}$ *and* $h \in \mathscr{H}$, *and*
3. *Terminal values* $r_N(f,0) \geq R_N(f,h) \geq r_N(f,h)$ *for all* $f \in \mathscr{F}$ *and* $h \in \mathscr{H}^+$.

In Assumption 3.2, the second condition implies that the worse the health state, the lower the NB; and the last condition implies that for the same fibrosis state, the terminal NB is highest in cancer-free state, followed by cancer state with treatment, and is lowest in cancer state without treatment. Throughout the paper, we use "decreasing" to refer to "non-increasing" for simplicity.

Assumption 3.3 *For cancer patients with small and medium-sized tumors, NB of immediate treatment is no worse than delaying the treatment until the next period. For cancer patients with large tumors, the total NB of immediate treatment is at least* $\frac{\kappa_1}{\zeta}$ *higher than that of delayed treatment. That is, for any* f *and* k, *(1)* $R_k(f,h) \geq W_k(f,h), h = 1,2$ *and (2)* $R_k(f,h) \geq W_k(f,h) + \frac{\kappa_1}{\zeta}, h = 3$.

In Assumption 3.3, the difference between the immediate and delayed treatment (i.e., $R_k(f,h) - W_k(f,h)$) can be interpreted as the opportunity reward of timely treatment for cancer patients. It is reasonable to assume that such opportunity reward should be non-negative as there is no clinical benefit for delaying liver cancer treatment. Especially for advanced stage liver cancer (i.e., $h = 3$), which is a fatal condition when left untreated, liver-disease related mortality is very high. As such, the opportunity reward is higher in such patients, and should be at least worth the screening cost per detected cancer case (κ_1/η) so that treatment benefit can offset the "average" screening cost.

We remark that *Screen* action is clearly not always the better action for maximizing the cost-effectiveness, because additional surveillance tests will increase cost but not necessarily health outcomes. Let $\pi_k(f)$ denote the decision for f-subpopulation in period k where the dependency of policy π on the initial state distribution is suppressed for simplicity of the notation. To examine the value of the additional surveillance test for a specific f-subpopulation at period t, for any given policy π, we can construct a pair of policies π^1 and π^2 such that π^1 has *Wait* action and π^2 has *Screen* action in period t for this given f-subpopulation, and actions in these two policies are the same elsewhere, i.e.,

$$\pi_t^1(f) = \mathsf{W}, \quad \pi_t^2(f) = \mathsf{E}, \quad \text{and } \pi_k^1(f') = \pi_k^2(f') = \pi_k(f') \text{ if } f' \neq f \text{ or } k \neq t. \tag{3.25}$$

We remark that the definitions of π^1 and π^2 depend on the specific time period t and specific fibrosis state f at which the two policies differ. However, again for simplicity of the notation, we drop these dependencies in the remainder of the analysis. Next, we define the *marginal surveillance benefit* for f-subpopulation at period t as

$$\Delta_t(f) = \sum_{h \in \mathscr{H}} b_t(f,h)[V_t^{\pi^2}(f,h) - V_t^{\pi^1}(f,h)], \tag{3.26}$$

which represents the expected incremental NB value by having a surveillance test compared with no screening in the given f-subpopulation at period t. In Proposition 3.1,

we provide a sufficient condition under which *Screen* is better than *Wait* action, i.e., an additional surveillance test improves the policy NB and thus the cost-effectiveness.

Proposition 3.1 *Let* π^1 *and* π^2 *be two identical policies except that the action is* Wait *in* π^1 *and* Screen *in* π^2 *for a specific* f*-subpopulation in period* t*, as defined in* (3.25)*. Then, policy* π^2 *is more cost-effective than policy* π^1 *(i.e.,* $\Delta_t(f) \geq 0$*), if*

$$\frac{\frac{1}{\zeta}\left[b_t(f,0)\kappa_0 + \left(\sum_{h\in\mathscr{H}^+} b_t(f,h)\right)\kappa_1\right] + \sum_{h\in\mathscr{H}^+} b_t(f,h)\Delta C_t(f,h)}{\sum_{h\in\mathscr{H}^+} b_t(f,h)\Delta Q_t(f,h)} \leq \lambda. \qquad (3.27)$$

A careful examination of (3.27) reveals that the left-hand-side indeed represents an upper bound on the ICER for performing an additional surveillance in f-subpopulation. In particular, the denominator represents the incremental QALYs for cancer patients by having immediate treatment compared with delaying treatment by one period, and the numerator represents the incremental costs, which consist of (1) expected surveillance cost per treated HCC case (therefore scaled by the factor $1/\zeta$) and (2) expected additional cost by immediate treatment. Thus, Proposition 3.1 can be interpreted as follows: cancer patients can achieve higher QALYs by timely treatment, which, however, is only possible by performing surveillance in all patients including those who actually are cancer-free and will not benefit from treatment. Considering such extra spending, if the overall ICER is still below the acceptable WTP value, performing a surveillance test is indeed beneficial and provides a good value for money. This condition further implies that the followings are the incentives for more frequent surveillance: 1) improved treatment NB, 2) improved sensitivity for the surveillance test, 3) increased cancer prevalence, and 4) reduced surveillance and diagnostic test costs.

Next, we introduce a definition for comparing transition matrices, which is needed in the following results.

Definition 3.2 *Let* $\mathbf{P} := [Q(j|i)]$ *and* $\mathbf{Q} := [Q(j|i)]$ *be two* $n \times n$ *transition matrices. We say that* $\mathbf{Q}$ *dominates* $\mathbf{P}$ *(or* $\mathbf{P}$ *is dominated by* $\mathbf{Q}$*), denoted as* $\mathbf{P} \preceq \mathbf{Q}$*, if* $\sum_{j \geq k} P(j|i) \leq \sum_{j \geq k} Q(j|i),\ \forall i, k = 1, \cdots, n.$

Hereinafter, we make the following additional assumptions for the following results.

Assumption 3.4 *Transition matrices* $\{\mathbf{F}_k, \mathbf{H}_k(f)\}$ *have following properties:*

1. *(Irreversible disease)* $F_k(f'|f) = 0$ *if* $f' < f$ *and* $H_k(h'|h, f) = 0$ *if* $h' < h$*,*
2. *(Accelerated progression)* $F_k(f'|f) \leq F_k(f'|f{+}1)$ *for* $f{+}1 \leq f'$ *and* $H_k(h'|h, f) \leq H_k(h'|h+1, f)$ *for* $h+1 \leq h'$*, and*
3. $\mathbf{H}_k(f) \preceq \mathbf{H}_k(f')$ *for* $f' \geq f$ *in any period* k*.*

In Assumption 3.4, the first condition implies that the disease progression under natural history is irreversible, i.e., patients cannot regress to a healthier state spontaneously without treatment. The second condition implies that patients in a worse health state are more likely to further progress to even worse states. We remark that the first two conditions imply a more general structure commonly assumed in medical decision making, known as the increasing failure rate (IFR) property (cf. (46, 18, 6, 243, 187)), which implies that disease progression accelerates as the disease condition worsens. Lastly, the third condition implies that tumor progression is more aggressive in patients with a more advanced fibrosis stage.

Assumption 3.5 *Sensitivity and specificity of surveillance test are greater than 50%.*

Assumption 3.6 *Let* $q'_k(f,h)$*,* $c'_k(f,h)$*, and* $\rho_k(f,h)$ *respectively represent the expected QALYs, costs, and survival probabilities under treatment for a detected cancer state* (f,h)*,* $f \in \mathscr{F}$*,* $h \in \mathscr{H}^+$ *in period* k*. At any period* k *and for any* f*-subpopulation:*

1. *Expected intermediate QALYs accumulated during treatment for small HCC is no higher than that of cancer-free state (i.e.,* $q'_k(f,1) \leq q_k((f,0),\mathsf{W})$*),*

2. *Expected intermediate cost of treatment for small HCC is no lower than expected intermediate cost in cancer-free state plus the test costs (i.e.,* $c'_k(f,1) \geq c_k((f,0),\mathsf{W}) + \kappa_0$*),*

3. *One period survival probability after treatment for small HCC is no higher than the probability that a cancer-free patient will remain in cancer-free state during the same period (i.e.,* $\rho_k(f,1) \leq P_k((f,0)|(f,0),\mathsf{W})$*).*

Assumption 3.6 implies that cancer-free state (i.e., $h = 0$) is better (in terms of QALYs, costs, and disease progression) than small HCC state (i.e., $h = 1$) even under treatment. Next, we present Proposition 3.2, where we show that for any policy π, the value function decreases as disease progression becomes more aggressive, which we use in proving Theorem 3.1.

Proposition 3.2 *For a given policy* π*, let* V_k^1 *and* V_k^2 *be the value functions associated with disease progression parameters* $\{\mathbf{F}_k^1, \mathbf{H}_k^1(f)\}$ *and* $\{\mathbf{F}_k^2, \mathbf{H}_k^2(f)\}$*, respectively. If* $\mathbf{F}_n^1 \preceq \mathbf{F}_n^2$ *and* $\mathbf{H}_n^1(f) \preceq \mathbf{H}_n^2(f)\ \forall f \in \mathscr{F}$ *and* $n \geq k$*, then* $V_k^1(f,h) \geq V_k^2(f,h), \forall f \in \mathscr{F}, h \in \overline{\mathscr{H}}$*.*

Next, in Theorem 3.1, we analyze the effect of disease progression on the value of surveillance test. We show that a surveillance test leads to a higher marginal benefit in population with a faster disease progression, which implies that surveillance should be more frequent in such a population. Given the heterogeneity in liver disease progression, such as accelerated fibrosis progression in patients infected with genotype 3 hepatitis C virus (26), this result provides practical insights about how a surveillance policy should be adjusted for certain subpopulation with a different disease progression rate.

Theorem 3.1 *As before, let* π^1 *and* π^2 *be two identical policies except that the action is* Wait *in* π^1 *and* Screen *in* π^2 *for* f*-subpopulation in period* t*. Let* $\Delta_t(f)$ *represent the marginal surveillance benefit for* f*-subpopulation in period* t*, as defined in* (3.26)*. Suppose two populations are associated with disease progressions* $\{\mathbf{F}_k^1, \mathbf{H}_k^1(f)\}$ *and* $\{\mathbf{F}_k^2, \mathbf{H}_k^2(f)\}$*, respectively, and corresponding marginal surveillance benefits* $\Delta_t^1(f)$ *and* $\Delta_t^2(f)$*, respectively. Given the same state distribution* $\{b_t(f,h)\}$ *in period* t*, if* $\mathbf{F}_n^1 \preceq \mathbf{F}_n^2$ *and* $\mathbf{H}_n^1(f) \preceq \mathbf{H}_n^2(f)\ \forall f \in \mathscr{F}$*, and* $n \geq t$*, then* $\Delta_t^1(f) \leq \Delta_t^2(f)$*,* $\forall f \in \mathscr{F}$*.*

The recent highly effective direct-acting antivirals (DAAs) have significantly changed the disease prognosis and the landscape in hepatitis C treatment. In that regard, it is interesting to analyze the effect of such advanced treatment options on the liver cancer surveillance. On one hand, these new DAAs effectively lower liver cancer risk, which, by Theorem 3.1, suggests less frequent HCC surveillance. On the other hand, DAAs also improve survival outcomes of HCC treatment, for example, by reducing the risk of HCC recurrence after liver transplant and thus improving the post-transplant survival (53). Hence, such improved treatment benefit motivates more frequent screenings. As a result, there is a trade-off and the net effect of improved hepatitis C treatment on optimal HCC surveillance policy is unclear.To examine this trade-off, in Theorem 3.2, we present a set of sufficient conditions which ensure that surveillance should be no more frequent as the risk of HCC decreases, while treatment outcomes improve.

Theorem 3.2 *For a given* f*-subpopulation, let* $\Delta_t^1(f)$ *and* $\{b_k^1\}$ *respectively represent the marginal surveillance benefit and state distribution for the base case, and* $\Delta_t^2(f)$ *and* $\{b_k^2\}$ *respectively represent the corresponding values for the case with reduced HCC risk and improved treatment outcome. Assume the same initial distributions* $b_1^1 = b_1^2$ *and that the total NB of HCC treatment increases by* $\Delta R_k(f,h)$ *for tumor*

state $h \in \mathscr{H}^+$ *and* $k \in \mathscr{N}$. *Then,* $\Delta_t^1(f) \geq \Delta_t^2(f)$ *if:*

$$\text{a)} \quad R_t(f,h) - R_t(f,h+1) \geq \big(r_t((f,h),\mathsf{W}) - r_t((f,h+1),\mathsf{W})\big) + \gamma H_t(h+1|h,f) R_{t+1}(f,h+1),\ h = 1,2,\ \textit{and} \tag{3.28}$$

$$\text{b)} \quad \max_{h \in \mathscr{H}^+} \{\Delta R_t(f,h)\} \leq \frac{\kappa_0 \left[b_t^2(f,0) - b_t^1(f,0)\right]}{\zeta \sum_{h \in \mathscr{H}^+} b_t^2(f,h)}. \tag{3.29}$$

Condition (3.28) implies that early stage HCC has higher treatment NB than a later stage HCC, and the difference, interpreted as the loss of treatment NB due to one-stage more advanced HCC, should be at least as high as the difference of intermediate NBs between the two state (h and $h+1$) plus the treatment NB after a potential progression to a higher state ($h+1$) in the next period. Condition (3.29) characterizes the trade-off between the treatment benefit for a small group (i.e., cancer patients) and the surveillance costs for the entire population, whether with cancer or not. In the right-hand-side of (3.29), the numerator $\kappa_0(b_t^2(f,0) - b_t^1(f,0))$ represents the incremental surveillance cost for cancer-free patients as the HCC risk is reduced, and the denominator represents the number of detected cancer patients—the small subgroup that will actually benefit from the surveillance. Then, Condition (3.29) implies that if the incremental treatment benefit is no higher than the incremental surveillance cost per detected cancer patient, there is no reason to perform more frequent surveillance.

3.5 Computational Study

In this section, we provide an overview of our parameter estimations, and present our numerical results for the optimal policies and policy comparisons. All computations are performed on a workstation with Intel Xeon 2.33 GHz processor and 12 GB RAM, and MIPs are solved using IBM ILOG CPLEX version 12.6 (Armonk, NY).

3.5.1 Parameter estimations

We estimate our model parameters using a variety of data sources, including the Hepatitis C Antiviral Long-term Treatment against Cirrhosis (HALT-C) trial data (144), the United Network of Organ Sharing (UNOS) data (168), a previously validated hepatitis C disease burden simulation model (HEP-SIM) that projects the changing prevalence of hepatitis C in the US (110), and several published studies. Model parameters, their base values, and corresponding sources are listed in Tables 3.2 and 3.3. Below, we provide an overview of our parameter estimations. Details of parameter estimations and model validations can be found in Appendix B.5.

Study population: In our base case analysis, we study a population starting surveillance at age 40. We do not consider patients below age 40 because the prevalence of both hepatitis C and HCC is very low before the age 40 (144). We used the HEP-SIM model (110) to estimate the initial distribution of fibrosis stages and HCC prevalence in the population (the HEP-SIM model did not consider natural history of HCC, health utility, and cost, so these parameters are estimated as described below).

Transition probabilities: Transition probabilities in our model depend on disease progression risks, mortality rates, and test accuracy values. Estimates of progression risks in non-tumor and tumor states are based on the best available evidence from the medical literature (217, 73, 144, 173). Excess mortality risks from disease state (f, h) due to liver disease are estimated from clinical observational studies (54, 91) and background mortalities for each age group are based on US life tables (15). Lastly, sensitivity/specificity of surveillance and diagnostic tests are estimated based on the results in the published literature (13, 182, 50).

We calibrate transition probabilities from cancer-free ($h = 0$) to small HCC state ($h = 1$) based on data from a large clinical trial, the HALT-C trial (144, 60). We remark that the available estimates of HCC incidences from the literature cannot be directly used in our model because these estimates are based only on detected HCC

Table 3.2: Model parameter values

Variables	Base value	Reference
Natural history parameters		
Annual progression risk in $\mathscr{F}$		
F3 to CC	0.116	(217)
CC to DC	0.039	(73)
Annual risk of developing small HCC		
From F3	0.008	Calibrated based on HALT-C trial
From CC	0.016	Calibrated based on HALT-C trial
From DC	0.078	Calibrated based on HALT-C trial, (173)
Monthly progression risk in $\mathscr{H}$		
Small to medium	0.056	(51)
Medium to large	0.036	(51)
Annual excess mortality		
CC	0.051	(54)
DC	0.265	(54)
Large HCC	0.750	(91, 13)
Mortality from other causes	life tables	(15)
Annual transplant probability for DC patients	0.023	(47)
Test accuracices		
Surveillance test: sensitivity	0.75	(13, 143)
Surveillance test: specificity	0.94	(203)
Diagnostic tests: sensitivity	0.90	(50, 182)
Diagnostic tests: specificity	1.00	Model assumption
Quality of life		
F3/CC/DC	0.93/0.90/0.80	(47, 49)
Age-adjustment factors		(97)
Age 20-29	0.924	
Age 30-39	0.911	
Age 40-49	0.880	
Age 50-59	0.854	
Age 60-69	0.831	
Age 70-79	0.793	
Age $\geq$80	0.765	
Costs		
Health state costs		(161, 47)
F3	\$1,394	
CC	\$1,626	
DC	\$18,064	
Screening costs		(13)
Surveillance test	\$185	
Diagnostic test	\$1,200	

CC, compensated cirrhosis; DC, decompensated cirrhosis; HCC, hepatocellular carcinoma.

cases, which are lower than the actual number of new HCC cases. In calibration, we use the existing incidence estimates as the base values for the corresponding transition

Table 3.3: Survivals, utilities, and costs of HCC treatment

Variables	Base value	Reference
Treatment effectiveness		
Liver transplantation		
Mean waiting time for transplant		
Within the Milan Criteria	158.7 days	UNOS data
Outside the Milan Criteria	205.6 days	UNOS data
1-year/5-year OS	0.887/0.738	(232)
Resection (5-year OS)		(105, 143)
Small HCC	0.66	
Medium HCC	0.52	
Large HCC	0.37	
Ablation (5-year OS)		
CC, small HCC	0.69	(140)
CC, medium HCC	0.61	(132)
DC, small /medium HCC	0.31	(132)
Palliative treatment (2-year OS)	0.41	(141)
Quality of life (first/subsequent years)		
Liver transplantation	0.69/0.73	(179)
Change in QOL after treatment		(220, 45)
Resection	86.8%/104.8%	
Ablation	88.4%/79.5%	
Palliative treatment	82.0%/68.6%	
Costs		
Liver transplantation (first/subsequent years)	$95,971/$25,208	(47)
Resection	$44,929	(125)
Ablation	$20,214	(125)
Palliative treatment	$18,539	(125)

OS, overall survival; HCC, hepatocellular carcinoma; CC, compensated cirrhosis; DC, decompensated cirrhosis; QOL, quality of life.

probabilities, and then increase the values until the percentages of detected HCC cases evaluated by the model (under similar surveillance strategies as in HALT-C trial) fall within 95% confidence interval of the estimates from the trial data.

Intermediate NB: To estimate intermediate QALYs, we adjust life years in each period by fibrosis stage- and age-specific quality of life (QOL), estimated from (49) and (97). Health-related costs associated with each fibrosis stage are estimated from (47) and costs of surveillance and diagnostic tests from (13). We adjust the intermediate NB by half-cycle correction to account for the fact that transitions among states occur continuously and not necessarily in the beginning or the end of a period (205).

Total NB of HCC treatment for each disease state is estimated by taking the weighted average of NBs for four major types of treatment: resection, liver transplantation, local ablation, and palliative treatment. For each treatment, total QALYs and costs are evaluated by a separate Markov model (details of which can be found in Appendix B.5.3), where survival rates and post-treatment QOLs are extracted from a series of published clinical studies (Table 3.3), and treatment costs are estimated from a population study using Medicare data (125). The weight for each treatment (i.e., the treatment distribution) is determined based on the best evidence available from US population-based studies (13, 70) and expert opinions.

Lastly, in our base case analysis, we consider a switching cycle of 4 years, and a WTP value of \$50,000/QALY to calculate corresponding NB based on QALYs and costs, which is in line with existing US-based cost-effectiveness studies in HCC surveillance (14, 13). We conduct extensive sensitivity analyses in Section 3.5.4 and Appendix B.4.

3.5.2 Optimal policies

In this section, we examine optimal surveillance policies and their implications. In particular, we seek answers to the following health policy questions: In which fibrosis stages is surveillance cost-effective? Should surveillance interval depend on fibrosis stage? What is the overall benefit of fibrosis stage-dependent surveillance policies? To what extent can dynamic surveillance policies improve cost-effectiveness compared with routine surveillance policies? Should the surveillance stop in older patients, and if so, what is the optimal age to terminate surveillance? We evaluate these questions by presenting the results for different classes of policies considered.

3.5.2.1 Fully dynamic policy

Figure 3.4 depicts the near-optimal fully dynamic policy (with 0.52% optimality gap after running the HS model for 16 hours). The dark area represents *Screen* action in each f-subpopulation, and white area represents *Wait* action. As expected, surveillance intervals are irregular throughout the planning horizon because we do not impose any particular structure on this policy. Therefore, such a policy could be impractical in real-world settings.

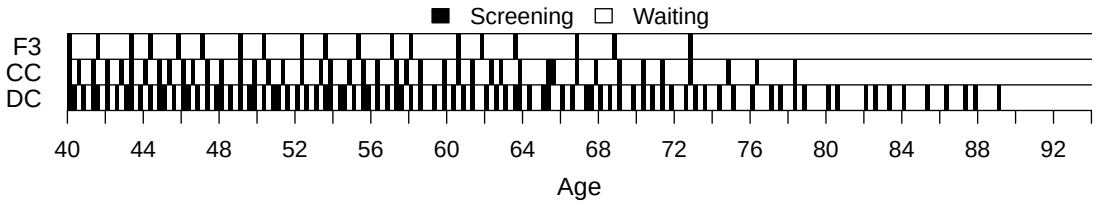


Figure 3.4: The optimal fully dynamic policy.

Note: Each row represents the policy for one subpopulation.

The optimal fully dynamic policy, however, provides several initial insights. First, we observe that HCC surveillance is recommended in F3 patients and would be cost-effective. This is an important clinical finding, because to our knowledge, the cost-effectiveness of surveillance in this patient cohort has not been evaluated before. Second, the fully dynamic policy suggests that surveillance intervals should not necessarily be identical in all patients. Instead, different f-subpopulations may need different surveillance frequencies. For example, surveillance is less frequent in early stages of hepatitis C (e.g., in F3 patients) and more frequent in advanced stages (e.g., in DC patients). Such a stratification by fibrosis stage is meaningful from a clinical perspective, because each f-subpopulation has distinctive clinical features (e.g., risk of cancer and disease progression, treatment eligibility and prognosis). Third, reducing the frequency of surveillance as patients age and ultimately terminating surveillance at a certain age (depending on the fibrosis stage) is more cost-effective

than continuing surveillance throughout patients' lifetimes.

3.5.2.2 M-switch policies

While the optimal fully dynamic policy provides useful insights, we acknowledge that it may not be practical for implementation. Therefore, we next focus on structured policies which have more practical appeal. In particular, we present the results of M-switch policies for various M values (0, 1, and 2) as described in Section 3.3.3.

We first present 0-switch (i.e., stratified) policies, where interval lengths are fixed (i.e., do not change with age) but could be different for each f-subpopulation (i.e., stratified by fibrosis stage). As presented in Figure 3.5, the optimal stratified policy recommends annual surveillance in F3 and CC subpopulations, and semiannual surveillance in DC subpopulation. Furthermore, the policy recommends terminating surveillance earlier in F3 subpopulation, compared with CC and DC subpopulations. These results further corroborate our findings from the fully dynamic policy that expanding HCC surveillance to F3 patients is cost-effective, but the surveillance should be less frequent and terminate earlier in F3 stage, compared with cirrhotic (i.e. CC and DC) stages.

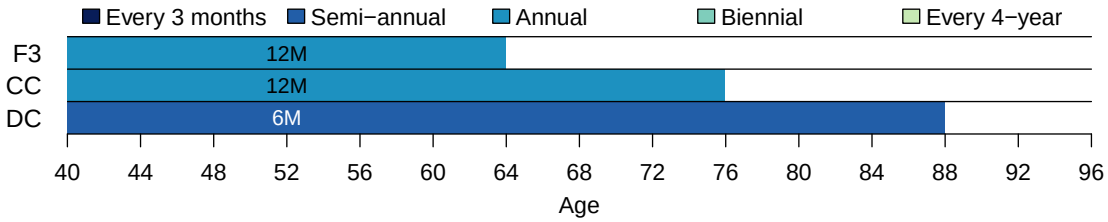
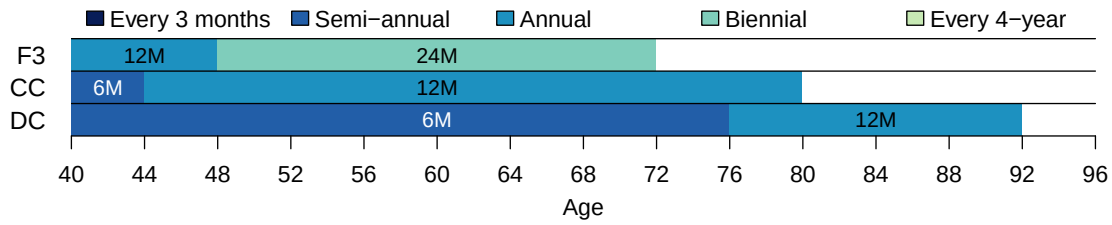


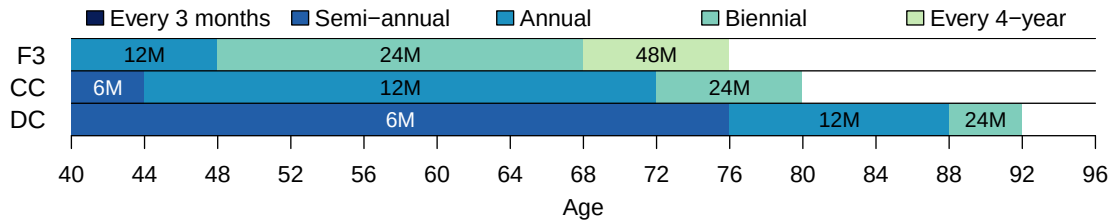
Figure 3.5: Optimal 0-switch (stratified) policy.

Note. The area shaded in different colors in each row represents different surveillance intervals over a certain age range for each subpopulation.

Figure 3.6 presents the optimal 1-switch and 2-switch policies. Compared with the 0-switch policy, these policies recommend more frequent surveillance at earlier ages. In older ages, 1-switch and 2-switch policies recommend less frequent surveillance but



(a) 1-switch policy



(b) 2-switch policy

Figure 3.6: Optimal M-switch policies.

extend surveillance to a higher terminating age. This aging effect is mainly driven by the decreasing life expectancy in the aging population. These findings also imply that when more flexibility is allowed in the policy structure, it is reasonable to allocate more clinical resources for prevention and treatment to younger patients, as early intervention achieves larger long-term treatment benefit in younger patients.

3.5.3 Policy comparisons

We compare the performance of our proposed surveillance policies against the currently recommended policies with respect to total expected net benefits, QALYs, LYs, costs, expected number of surveillance tests, number of HCC detected at each stage, and number of deaths due to liver disease. We further examine the incremental cost-effectiveness of each policy. In evaluating routine policies, we first assume that only cirrhotic patients undergo surveillance tests, as this is the current common clinical practice (33, 143). Later in Section 3.5.3.2, we also analyze the effect of expanding surveillance to F3 patients in routine policies.

3.5.3.1 Comparisons of the M-switch policies

Table 3.4 compares the performances of M-switch policies against no surveillance and fully dynamic policy. Among all policies considered, as expected, the fully dynamic policy is the most cost-effective policy. While we acknowledge that such a policy is impractical, we remark that the results from the optimal fully dynamic policy provide the maximum NB that can be achieved, which can be used as a benchmark metric for interpreting our results.

Table 3.4: Policy comparisons

Surveillance policy	NB	QALY	LY	Cost	# Tests	# per 1,000 patients			
						# HCC detected			DLD
						Small	Medium	Large	
No surveillance	336,525	7.398	9.492	33,353	0.0	0	0	0	669
Fully dynamic	347,529	7.794	10.137	42,160	11.6	127	39	7	515
M-switch									
0-switch	347,352	7.782	10.119	41,759	10.0	120	42	8	518
1-switch	347,425	7.791	10.133	42,139	11.7	125	39	7	516
2-switch	347,426	7.791	10.133	42,125	11.7	125	39	7	516

NB, net benefit; QALY, quality-adjusted life year; LY, life year; HCC, hepatocellular carcinoma; DLD, death due to liver disease.

Interestingly, our results show that the NBs of optimal M-switch policies are very close to that of the optimal fully dynamic policy (Table 3.4). Further, with respect to all other measures, optimal M-switch policies perform similarly to the fully dynamic policy. Another interesting finding is that having more than one switch in surveillance intervals does not substantially improve the outcomes (results of policies with more than two switches are presented in Appendix B.4.1). This finding implies that 1-switch policy indeed is clinically very appealing: it achieves desirable performance close to the optimal fully dynamic policy; yet, unlike the fully dynamic policy, it has a simple structure and hence is practical for implementation. Therefore, in the remainder of our analysis, we do not consider policies with more than one switch.

3.5.3.2 Comparisons with currently recommended policies

Table 3.5: Routine policy comparisons for cirrhotic patients.

Surveillance policy	NB	QALY	LY	Cost	# Tests	# per 1,000 patients			
						# HCC detected			DLD
						Small	Medium	Large	
No surveillance	336,525	7.398	9.492	33,353	0.0	0	0	0	669
Routine policies (cirrhotic patients)									
24-month	344,470	7.661	9.923	38,573	3.7	65	42	12	562
12-month	346,323	7.731	10.037	40,225	7.0	96	42	9	537
6-month*	346,538	7.773	10.106	42,133	13.6	125	33	5	524
3-month	344,331	7.794	10.140	45,391	26.6	147	21	3	517

NB, net benefit; QALY, quality-adjusted life year; LY, life year; HCC, hepatocellular cancer; DLD, death due to liver disease; *: currently recommended policy.

Most current HCC management guidelines recommend every 6-month surveillance in cirrhotic (i.e., CC and DC) patients, and remain uncertain about the surveillance frequency in non-cirrhotic (F3) patients (33, 143). Compared with the recommended 6-month surveillance for cirrhotic patients (which appears as the most cost-effective routine policy when surveillance is offered to only cirrhotic patients), M-switch policies recommend fewer number of surveillance tests while improving health outcomes (see Tables 3.4 and 3.5). For example, compared with the 6-month surveillance policy, 1-switch policy saves about 1.9 tests per person. Given that there are about 82,000 hepatitis C patients at risk of HCC in the 40-45 year-old group (110), this leads to 156,000 fewer tests only for this age group. This finding implies that M-switch policies can utilize limited surveillance and treatment resources in a more efficient way, and improve cost-effectiveness of HCC surveillance.

We next examine the effect of expanding surveillance to F3 subpopulation in routine surveillance policies (Table 3.6). This analysis reveals three important findings. First, expanding surveillance to F3 patients improves the cost-effectiveness in most policies (compare NB in Tables 3.5 and 3.6). However, this is not always true. In

Table 3.6: Routine policy comparisons when surveillance is expanded to F3 patients.

Surveillance policy	NB	QALY	LY	Cost	# Tests	# per 1,000 patients			
						# HCC detected			DLD
						Small	Medium	Large	
Routine policies (F3+cirrhotic patients)									
24-month	345,348	7.694	9.975	39,348	5.0	75	45	12	550
12-month	347,131	7.768	10.095	41,264	9.4	109	44	9	524
6-month	346,912	7.811	10.166	43,650	18.0	141	33	4	511
3-month	343,662	7.832	10.200	47,933	35.3	165	20	1	505

NB, net benefit; QALY, quality-adjusted life year; LY, life year; HCC, hepatocellular cancer; DLD, death due to liver disease.

particular, with the expanded surveillance, excessive surveillance such as every 3-month policy becomes less cost-effective. This is because in F3 patients have much lower risks of HCC, and thus, the limited benefits by surveillance cannot offset the costs of excessive surveillance. Second, when surveillance is expanded to F3 patients, among the routine policies, every 12-month (rather than every 6-month) surveillance becomes the most cost-effective policy. Third, expanding surveillance to F3 patients in routine policies is not sufficient to achieve a similar level of NB compared with M-switch policies (compare Tables 3.4 and 3.6). This further corroborates our findings that M-switch policies can lead to more efficient use of health care resources.

3.5.3.3 Incremental cost-effectiveness analysis

Incremental cost-effectiveness analysis can help decision-makers determine if they are willing to pay additional cost for improved health outcomes. We analyze the incremental cost-effectiveness of the policies considered by following a standard approach used in health economic evaluation (85, 64). In particular, we first order policies from the least to the most effective (i.e., from the lowest to the highest QALYs), and eliminate the dominated policies (i.e., policies with lower QALYs but higher costs). Then, for each successive non-dominated policy, we calculate ICER by dividing the incremental costs by the incremental QALYs.

Table 3.7 presents ICERs for the surveillance policies considered and Figure 3.7

Table 3.7: Incremental cost-effectiveness ratios (ICERs).

Policy	QALY	Cost ($)	ICER=ΔCost/ΔQALY
No surveillance	7.398	33,353	-
24-month (cirrhotic)	7.661	38,573	19,826
24-month (F3+cirrhotic)	7.694	39,348	23,450
12-month (cirrhotic)	7.731	40,225	23,666
12-month (F3+cirrhotic)	7.768	41,264	28,126
6-month (cirrhotic)	7.773	42,133	(dominated)
0-switch	**7.782**	**41,759**	**34,594**
1-switch	**7.791**	**42,139**	**41,936**
3-month (cirrhotic)	7.794	45,391	(dominated)
6-month (F3+cirrhotic)	7.811	43,650	75,673
3-month (F3+cirrhotic)	7.832	47,933	207,436

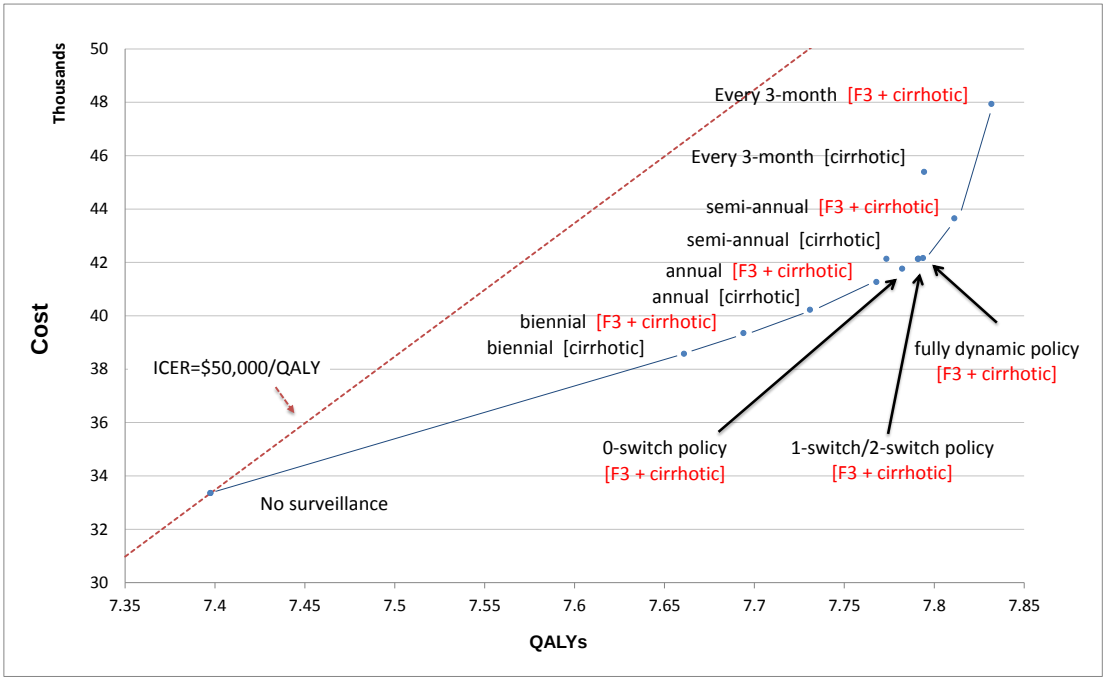


Figure 3.7: Cost and effectiveness of surveillance policies.

presents the corresponding efficiency frontier. As shown in Table 3.7, 0-switch and 1-switch policies have acceptable ICERs well below the WTP threshold of $50,000/QALY. Compared with M-switch policies, routine policies are either less effective with lower QALYs or are more effective but too costly with unacceptable ICERs (i.e. ICER

>$50,000/QALY). As shown in Figure 3.7, M-switch policies are on the efficiency frontier and are more effective (i.e., have higher QALYs) than other policies with ICERs below the baseline WTP. Every 3- and 6-month surveillance in F3 and cirrhotic patients can further improve the health outcome but at substantially higher costs, leading to unacceptable ICERs.

3.5.4 Robustness checks and probabilistic sensitivity analysis

We conduct a series of sensitivity analyses to assess the robustness of our results. In the following, we present the probabilistic sensitivity analysis (PSA), and more extended analyses can be found in Appendices B.4.2 and B.4.3.

In the PSA, we account for uncertainty in model inputs by randomly sampling the value of all parameters simultaneously using the recommended statistical distributions (31). For each distribution, we assume the coefficient of variation to be 10%. We repeat the sampling 10,000 times, and for each sampled set of parameters we evaluate the policies and calculate ICERs.

Figure 3.8 presents the cost-effectiveness acceptability curves for the optimal stratified and 1-switch policies, which represent the probability that each policy is cost-effective (i.e., with ICERs below a given WTP value) in comparison with the current common practice, i.e., every 6-month surveillance in cirrhotic patients. These results show that 0-switch and 1-switch policies are cost-effective with high confidence, at least 87% and 97% at the baseline WTP value of $50,000/QALY, respectively. Furthermore, they are also cost-saving (i.e., have higher QALYs with less cost) in about 62% and 37% of PSA samples, respectively (denoted by vertical intercepts in Figure 3.8). Thus, PSA implies that our model results are fairly robust, which further supports our findings in the cost-effectiveness of M-switch policies.

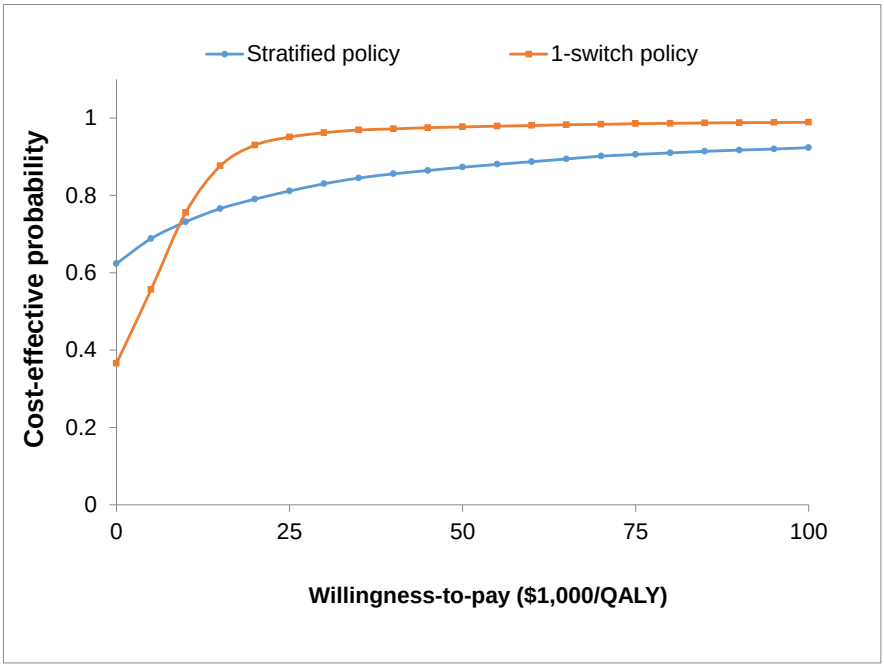


Figure 3.8: Cost-effectiveness acceptability curves (CEACs) for 0-switch and 1-switch policies compared withevery 6-month surveillance in cirrhotic patients, the current common practice.

Note. A CEAC represents the probability that the estimated ICERs are below the given WTP value.

3.6 Discussion

With the rising burden of liver cancer in the US, there is an imperative need to diagnose HCC in early stages by regular surveillance in high-risk patients. Even though HCC surveillance has a potential to reduce the disease burden, its current practice is substandard because of several reasons including lack of clinical trials and confusion over the optimal frequency. Under such situations, mathematical models can provide important clinical insights that otherwise are not feasible to obtain. In this work, we develop a general modeling framework to systematically evaluate the health and economic benefits of HCC surveillance policies in hepatitis C patients.

Unlike previous sequential medical decision making models, our proposed modeling framework permits explicit formulation of structured policies (i.e., M-switch policies) that achieve high performance but are at the same time also practical for implementation. Our modeling framework also effectively captures the condition that the disease progression is jointly characterized by observable and unobservable states, but intervention policies explicitly depend on only observable states. Although in this study we focus on HCC surveillance, our modeling framework is general and could be applied to many other diseases with or without precursor/comorbid conditions. For instance, our modeling framework could be applied to coronary disease surveillance with electron-beam CT which can depend on patients' hypertension status or other risk factors (90); breast cancer screening with mammogram or other imaging techniques which can be tailored by patients' pre-diabetic/diabetic state (126); and esophageal cancer surveillance with endoscopy which can be adjusted based on patients' precancerous dysplasia conditions (106).

Our findings unravel the complex links between disease progression, mortality, test accuracy, health outcomes, costs, and frequency of surveillance policies. First, unlike the current guidelines, we find that surveillance should be stratified by fibrosis stage. Second, we find that expanding surveillance to F3 patients improves the overall cost-effectiveness of surveillance; however, F3 and CC patients should have less frequent surveillance than the currently recommended every 6-month policy. This is an important finding filling the gap in the literature about the cost-effectiveness of HCC surveillance for F3 patients. Third, we find that surveillance should become less frequent as patients age, and terminate after a certain age.

Our study has limitations. Complex model parameterization is based on published estimates from multiple sources and different study cohorts. We also calibrate several parameters to the HALT-C trial cohort. Due to lack of population-based estimates, our model parameters may suffer from potential biases by specific study cohort. To

assess the impact of parameters on model results and ameliorate this potential bias, we conduct extensive probabilistic sensitivity analyses which have shown that the cost-effectiveness results of our base case policies are robust against parameter variabilities.

This study can be extended in several ways. One is to incorporate other screening modalities in HCC surveillance, such as CT and MRI. They are more accurate tests but are also more costly. Screening with alternate modalities has been proposed to improve the accuracy of surveillance, but the cost-effectiveness of such strategies has not been addressed yet. In addition, we may investigate the effect of low utilization and adherence in practice on the outcomes of surveillance (204, 201). Patients may be less likely to follow very frequent surveillance (e.g., every 3-month screening), and thus the structure of optimal policy can be different. Lastly, given the recent availability of highly-effective hepatitis C treatments, the landscape of HCC surveillance could change in future. In that regard, our modeling framework can be used to evaluate the impact of new hepatitis C treatments on HCC surveillance.

CHAPTER IV

OPTIMAL RESOURCE ALLOCATION FOR HEPATITIS C ELIMINATION

4.1 Introduction

The burden of hepatitis C virus (HCV) infection continues to rise globally. More than 185 million people, 2.8% of the world population, are infected with hepatitis C (165). HCV infection can cause a series of life-threatening complications, such as cirrhosis, liver failure, and liver cancer. In the US, the mortality rate due to HCV has already surpassed that due to HIV infections (145).

In the last few years, significant breakthrough has been made in hepatitis C treatment. The modern antiviral agents can achieve cure rates higher than 90% within only 8-12 weeks, compared with the previous regimens with low cure rates of only 60%, long treatment duration, and significant side effects (189). The advent of highly effective treatments not only makes it possible to cure hepatitis C in individuals, but also shows the potential to eliminate the disease in a population. In other words, it is possible to reduce the incidence of new infections (close) to zero in a defined geographic area through deliberate intervention efforts (63).

Recently, several efforts and global movement have been dedicated to preventing, diagnosing, treating, and developing national programs to ultimately eliminate viral hepatitis as a public health problem. The World Health Assembly recently adopted the first "Global Health Sector Strategy on Viral Hepatitis, 2016-2021" developed by the World Health Organization (WHO, (169)). This strategy sets the elimination target for hepatitis C by the year 2030: to reduce new infections by 90%, and to reduce deaths due to viral hepatitis by 65%, as well as other targets for screening and

treatment coverage. With the global momentum of hepatitis C elimination, many countries have started taking actions to address the epidemic of hepatitis C, especially in those places with high disease prevalence such as Egypt, Georgia, Mongolia, and certain regions of India like Punjab.

However, meeting HCV elimination targets faces several challenges. First, many infected individuals are not aware of their infection status, which could become the bottleneck for achieving the disease elimination—availability of efficacious HCV treatment won't be of much use if patients are not aware of their infection. In addition, unawareness of HCV infection also fuels ongoing HCV transmission. Therefore, it is critical to identify the infected patients who can then be treated with available therapies.

Second, limited resources and budget for scaling-up screening and treatment to eliminate HCV can become a barrier. Both screening and treatment at a population level could be costly and unaffordable. Unfortunately, the current level of funding for HCV interventions is severely insufficient. For example, while HCV prevalence is 3-4 times higher than that of HIV, the appropriation for HCV is only less than 2% of resources for HIV control (66). Hence, in such a situation with limited resources, an efficient resource allocation strategy becomes even more critical.

In this chapter, we study a resource allocation problem for hepatitis C elimination. In particular, we consider two types of intervention, *screening* and *treatment*. The objective is to determine the optimal screening and treatment rates over the next 15 years (2015-2030) in a target population. We consider two different model setups: (1) minimize the total cost of screening and treatment to achieve certain epidemic reductions in line with the WHO targets; and (2) minimize the disease burden, the cumulative number of infected patients, subject to a constant budget rate.

Our contributions of this chapter are twofold. From a theoretical viewpoint, we show the asymptotic stability behavior of the hepatitis C transmission model, and

characterize the structure of the optimal policies for two optimal control problems which are nonlinear, non-convex, and may suffer from singular solutions. From a practical viewpoint, both analytical and numerical results lead to important and practical policy implications. In particular, we find that for the cost minimization problem, a simple policy structure with a constant screening rate represents a good balance between model tractability, policy performance and practicality. Among such a class of policies, a minimal constant screening rate with 100% (or maximum) treatment rate is found optimal. Lastly, for the fixed budget allocation problem, we find that the optimal policy is to prioritize treatment first, and then to allocate the remaining budget, if any left, for screening.

The remainder of this section is organized as follows. In Section 4.2, we review the relevant literature in infectious disease modeling and the resource allocation problems in epidemic control. In Section 4.3, we introduce a system dynamic model for hepatitis C transmission, and theoretically analyze the reproduction number and the equilibrium stability. In Sections 4.4 and 4.5, we respectively present the two optimal control problems: (1) minimize the total screening and treatment cost subject to the final prevalence target constraint, and (2) minimize total disease burden subject to the budget constraint. For each problem, we examine the structure of the optimal policy, conduct extensive numerical analyses, and identify the policy implications from both analytical and numerical results. Finally, we discuss the limitations and future extensions of this study and conclude in Section 4.6.

4.2 Literature Review

We categorize the relevant literature to our study into the following two groups. First, we review the classic infectious disease modeling and the mathematical epidemiology literature, and particularly existing studies on HCV transmission models. Second, we review the resource allocation optimization problems in the context of epidemic

control.

4.2.1 Infectious disease modeling

The literature of mathematical epidemiology have experienced a rapid growth starting from the middle of 20th century. Many compartment models have been developed to analyze various aspects of infectious disease such as vertical transmission, mixing population, vaccination, quarantine, treatment (29). The set of compartments depends on the characteristics of the disease and the study objective, which may include, but is not limited to, class M with passive immunity, susceptible class S, exposed class E, infective class I, and recovered class R. These compartments can be assembled in many ways to capture the key dynamics of the problem, such as the classic SI, SIS, SIR, and SIRS models, and the more extended versions like SEIRS, MSEIR models (102). These modeling studies typically focus on the analysis of the basic reproduction number (101, 58, 59), local and global stability analysis of disease-free equilibrium and epidemic equilibrium (162, 122, 226). Excellent reviews of mathematical epidemiology models and analyses can be found in (9, 152).

A number of mathematical models have been developed for HCV transmission in the dynamic modeling literature (52). Martin and colleagues develop a compartment model of HCV transmission among injection drug users (IDUs) to assess the effect of antiviral therapies on the prevalence reduction in the UK setting (155). The model is later extended by incorporating the status of risk reduction intervention among IDUs in addition to their HCV infection states. They find that scaling up the risk reduction intervention results in modest prevalence reduction (227), whereas scaling-up treatment could have a major preventive impact (156). The model is also used to examine the joint impact of treatment and risk reduction interventions on the prevalence among IDUs (153). Echevarria et al. modify the model for IDUs in metropolitan Chicago and analyze the effect of scaling-up treatment among such

population (65). Elabasha explicitly models the reinfection after previous successful treatment as separate compartments (71), shows local and global stability results of the model equilibrium, and presents the numerical results for IDU population in the US.

As another class of modeling method, simulation modelings exhibit greater flexibility in capturing finer details in disease dynamics. For example, in addition to the dynamics of disease transmission at a population level, a simulation model can also capture the progression of disease stages at an individual patient level. Hahn and colleagues construct a stochastic individual-based simulation model based on injecting behavior data collected from young IDUs in San Francisco to project the short- and long-term effects of potential HCV vaccine (94). To evaluate the impact of imperfect follow-up during the HCV cascade of care, Linas and colleagues develop a state-transition simulation model of disease progression and care delivery, and use it to identify the intervention strategies to be prioritized along the cascade of care (138). He et al. develop an agent-based simulation to model the transmission and progression of HCV inside and outside the US prison, and find that universal screening in prisons is highly cost-effective and can also reduces disease burden in the outside community (100).

4.2.2 Resource allocation and optimal control in epidemics

Resource allocation problems for epidemic control have been studied for many infectious diseases, and most extensively for HIV control in a variety of settings. Alistar et al. minimizes the reproduction number subject to a budget constraint to optimize the budget allocation of prevention and treatment for HIV epidemic control (8). Kok et al. determine the optimal allocation of HIV test among targeted and routine programs to minimize the new infections in a dynamic model (120). An optimal portfolio problem raises when there exist multiple intervention options with

different effects and costs. For a given budget, the optimal intervention portfolio is determined for interacting populations (240, 7), for multiple independent populations (28), and when the combination is allowed to change over the planning horizon (241). For multi-level resource allocation problems which is commonly encountered in HIV prevention practice, Lasry et al. solve the optimal budget allocation at two levels and compare with equity-based heuristic allocation policies (127), and Malvankar-Mehta and Xie identify the optimal incentives provided to decision makers at different levels to maximize the total infections averted (149). Deo et al. formulate a planning problem for screening and care of HIV as a non-linear mixed integer program, and solve the optimal screening rate and the staffing level subject to budget and capacity constraints using integer programming (56). Khademi et al. develop a Markov decision process (MDP) formulation for the optimal allocation of scarce HIV treatment resources which allows early discontinuation of treatment, and employ the approximate dynamic programming approach to lower bound the price of nonabandonment by comparing with the standard life-long treatment policy (116).

Optimal control theory, which has been extensively used in engineering and economic applications (118, 190, 191, 82), is also a commonly utilized methodology in analyzing optimal epidemic control problems. The Pontrayagin maximum principle provides a set of necessary optimality conditions for optimal control problems, which can be used to obtain analytical solutions of the optimal policies in some cases. Optimal treatment policy is studied for the spread of crop disease over regional scale in a contact-process model (80). Rowthorn and Toxvaerd fully characterize the optimal prevention and treatment policy in a classic SIS epidemic model (183). There exists abundant body of literature on the optimal intervention policies in classic epidemic models (e.g., SIR model), such as vaccination plus treatment (129), pulse vaccination (195), intervention with time delay (239, 30), and educational campaign plus treatment (22). In particular, Martin et al. examine the optimal hepatitis C treatment

policy in IDU population with respect to various objectives and budget constraints (154). Hansen and Day show the structure of the optimal prioritization of vaccination and isolation in an SIR model with budget constraints (98). Optimal allocation of limited total budget for two coupled subpopulations are analyzed in SIRS model (160), and in SIS model with quarantine (184) or heterogeneous populations symmetric and asymmetric transmission (159).

In addition to infectious disease control problems, dynamic modeling and optimal control theory have also been applied in a broad set of health care problems, such as liver transplant prioritization policy (5), dialysis delivery design subject to budget and capacity constraints (242), and cancer treatment design (128, 62, 109).

4.3 Disease Transmission Model

We consider the HCV transmission dynamics in an open population with new births. The population are partitioned into four compartments by their hepatitis C infection status (Figure 4.1): (1) Compartment S for the susceptibles, who have never been infected with HCV, (2) Compartment I for the infected but undiagnosed, who are currently infected with HCV but unaware of their infection, (3) Compartment D for the diagnosed infections, who are aware of their infection awaiting treatment, and (4) Compartment R for the recovered, who are cured (i.e., sustainable virologic response [SVR] achieved) by successful treatment. The susceptible population (i.e., compartment S) can become infected by contacting with infected patients including both undiagnosed and diagnosed cases (i.e., compartment I and D). Infected patients can be diagnosed only through HCV screenings. Obviously, only diagnosed patients are eligible for treatment. Those who failed treatment remain in compartment D and can be retreated later. Although current direct antiviral agents have high cure rate, recovered patients are still at risk of reinfection, especially in high risk population like IDUs.

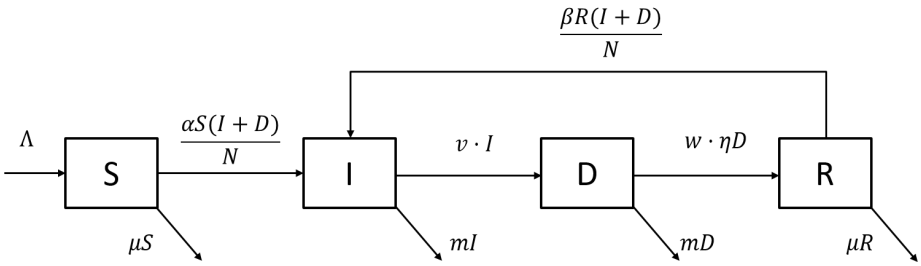


Figure 4.1: Dynamic model for HCV transmission.

We define the state variable $x(t) = (S(t), I(t), D(t), R(t))$ to represent the population size of each corresponding compartment at time t, and $N(t) = S(t) + I(t) + D(t) + R(t)$ to represent the size of the total population at time t. While state variable depends on time t in principle, we suppress the dependency of time t in the notation for simplicity, unless absolutely necessary.

We consider the *standard incidence* $\frac{\alpha S(I+D)}{N}$ for new infections and $\frac{\beta R(I+D)}{N}$ for reinfections, where α and β represent the infection rate for the susceptible population and the treated patients, respectively. The rationale of standard incidence formulation is that an uninfected individual has limited contact with the infected population, and thus the likelihood of being infected (i.e., the *infection force* defined as $\frac{\alpha(I+D)}{N}$ and $\frac{\beta(I+D)}{N}$, respectively) is proportional to the fraction of the infected population in the population. The uninfected population (compartments S and R) have a per capita death rate of the background mortality μ, and the infected population (compartments I and D) have higher mortality rate $m > \mu$ to represent the excess liver-related mortality due to the infection. We assume a constant total birth rate Λ of the susceptible population.

Remark 4.1 *In the epidemic modeling literature, another commonly used incidence formulation is the* mass action incidence*, defined as $\alpha S(I+D)$ and $\beta R(I+D)$ accordingly in our model setup, which, however, is deemed not suitable for modeling HCV transmission. This is because it has an infection force proportional to the absolute number of infections, implying that the contact between the uninfected people and the*

infected population can increase indefinitely as the infected population grow. On the other hand, it could be more suitable for studying infectious diseases like influenza and SARS (152).

We consider two different types of interventions: *screening* and *treatment* for hepatitis C. We denote the control variable $u(t) = (v(t), w(t))$, $v(t), u(t) \in [0, 1]$ to represent screening rate and treatment rate at time t, respectively. Screening is applied to all population except the diagnosed patients in compartment D, whereas treatment is only available to diagnosed patients. Since currently available screening tests for hepatitis C are highly accurate with the sensitivity and specificity close to 99% (194), we assume perfect screening test, and thus the transition rate from compartment I to D is simply the screening rate v. We denote η as the SVR rate of treatment, which implies the transition rate to be $w\eta$ from compartment D to R. Mathematically, the above-described population dynamics are presented using a set of nonlinear ordinary differential equations as follows. We use the notation $\dot{x} := \frac{dx}{dt}$ to represent the time derivative of variable x. Then,

$$\dot{S} = \Lambda - \alpha S \frac{I + D}{N} - \mu S \tag{4.1}$$

$$\dot{I} = (\alpha S + \beta R) \frac{I + D}{N} - mI - vI \tag{4.2}$$

$$\dot{D} = vI - mD - w\eta D \tag{4.3}$$

$$\dot{R} = w\eta D - \beta R \frac{I + D}{N} - \mu R, \tag{4.4}$$

$$N = S + I + D + R. \tag{4.5}$$

The initial condition is determined by a given population profile (S_0, I_0, D_0, R_0):

$$(S(0), I(0), D(0), R(0)) = (S_0, I_0, D_0, R_0). \tag{4.6}$$

All variables and parameters are summarized in Table 4.1.

Table 4.1: List of variables and parameters.

	State variables: $x(t) = (S(t), I(t), D(t), R(t))$
$S(t)$	Number of suspectible (uninfected) persons
$I(t)$	Number of infected but undiagnosed patients
$D(t)$	Number of infected and diagnosed patients, waiting for treatment
$R(t)$	Number of recovered patients (cured from previous hepatitis C treatment), but at risk of reinfection
	Control variables: $u(t) = (v(t), w(t))$
$v(t)$	Screening rate in S, I, and R population
$w(t)$	Treatment rate in D population
	Transmission model parameters:
Λ	Birth (Arrival) rate of the population
α	Infection rate for new infections in susceptible population
β	Infection rate for reinfections in treated patients
μ	Background mortality
m	Mortality rate of the infected population (including the liver-related mortality)
η	Sustainable virologic response (SVR) rate of hepatitis C treatment
	Cost parameters:
c_s^a	Cost of HCV antibody screening test
c_s^r	Cost of HCV RNA test
c_d	Cost of antiviral treatment
B	Budget for screening and treatment per year

It is safe to assume that reinfection rate should be no higher than that of new infections, as a study has found that persons who have previously cleared HCV have lower odds of developing infections (89) (also see the parameter estimation in Section 4.4.3). Also, it is reasonable to assume that the infection rate α is higher than mortality rate m for the infected population. This is because when there is no intervention, $\dot{I} + \dot{D} = \left(\frac{(\alpha S + \beta R)}{N} - m\right)(I + D)$; if $\alpha < m$, then $\dot{I} + \dot{D} < 0$. That is, in this case, the disease transmission is so slow that the disease is not self-sustaining and will extinct even without any extra intervention efforts, which is not true in the context of HCV epidemics. We formally state these assumptions below.

Assumption 4.1 $\alpha \geq \beta$, $\alpha > m$.

Lastly, we remark on the particular choice of the compartment model for this problem in comparison with the model formulated in Chapter 3. Although both of Chapter 3 and this chapter are related to the disease modeling for hepatitis C, the research questions are at different levels and thus the models differ by their definitions of states or the compartments. In particular, Chapter 3 focuses on the cancer screening policy for the population who have already infected with hepatitis C. Since the stages of infection affect the disease progression dynamics (e.g., changing cancer risks, mortality risks) and subsequently influence the screening policy, a model that explicitly captures the disease stages is deemed necessary.

In this chapter, the intervention policies target at a broader population including not only hepatitis C-infected but also uninfected population. The main focus is to study the intervention policy that controls disease transmission in the population. Thus, a model is deemed necessary if it at least distinguishes the subgroups of the population by their infection status. We choose a minimal set of compartments as defined in (4.1)-(4.4) to reflect the major differences of infection status in the population. Since intervention policy will not be tailored by disease stages in our basic problem setting, we do not further stratify the compartments by the stages of hepatitis C infections. The simple representation of the dynamic system also makes it analytical tractable to derive policy structures from the optimal control formulations as presented in the following sections.

We acknowledge the limitation that the compartment model (4.1)-(4.4) cannot capture the exact natural history because of the aggregation of the disease stages. For example, it has only one overall mortality m for infected population (in compartments I and D), which ignores the differences in the liver-related mortality by specific stage of HCV infection: the excess mortality is negligible for asymptomatic patients in early stage (i.e., F0-F3 fibrosis stages), but is very high for patients with

advanced liver disease (e.g., F4 stage fibrosis, decompensated cirrhosis, liver cancer). However, since the majority of the population in our model are uninfected and newly infected population, such a simplification is unlikely to significantly skew the population dynamics and alter the policy implications.

4.3.1 Reproduction number and stability analysis

In the following theoretical analyses of the HCV transmission model (4.1)-(4.4), we will calculate the reproduction number $\mathscr{R}_c$ and examine asymptotic behavior of the epidemic dynamics.

Before the analyses, we first check the boundedness of the system. Define the feasible set as

$$\mathscr{X} := \{(S, I, D, R) \in \mathbb{R}^4_{\geq 0} : N(t) \leq \max\{\frac{\Lambda}{\mu}, N_0\}\}. \tag{4.7}$$

Proposition 4.1 *The closed set $\mathscr{X}$ is* positively invariant *with respect to the model* (4.1)-(4.4). *That is, all solutions with initial conditions in set $\mathscr{X}$ will remain in $\mathscr{X}$ for any admissible control $u(t)$ for $t \geq 0$, i.e., $x_0 \in \mathscr{X} \implies x(t) \in \mathscr{X}$, $\forall t > 0$.*

Proposition 4.1 implies that the system (4.1)-(4.4) is always bounded for any given initial population N_0. Therefore, no particular state space will be excluded from the following qualitative analyses.

In the infectious disease modeling literature, the basic reproduction number $\mathscr{R}_0$ is a fundamental measure, which quantifies the severity of an epidemic, and is defined as the average number of new secondary infections caused by an index case during his life time when introduced to a susceptible population, *in the absence of any prevention or intervention* (58). In our analyses, we are interested in the effects of interventions on the epidemics, and thus our analyses instead focus on a counterpart of $\mathscr{R}_0$ in the presence of intervention, the *control reproduction number* denoted by $\mathscr{R}_c$. For simplicity, we refer to the control reproduction number $\mathscr{R}_c$ as the reproduction number throughout the remainder of this chapter.

Proposition 4.2 *The reproduction number $\mathscr{R}_c$ for the HCV transmission model* (4.1)-(4.4) *is given by:*

$$\mathscr{R}_c(v,w) = \frac{\alpha(v+\eta w+m)}{(v+m)(\eta w+m)} \tag{4.8}$$

The mathematical representation of the reproduction number $\mathscr{R}_c$ is indeed in line with its definition. Consider the following process. One infected individual—the index case—starts in compartment I and stays in compartment I for an average time of $\frac{1}{v+m}$; during this period of time, he will infect $\frac{\alpha S(I+D)}{N} = \alpha$ persons in the susceptible population per time unit (note that $\frac{S}{N} = 1$ and $I = 1$ as this infected individual enters a population with all susceptible population). Then he moves to compartment D with probability $\frac{v}{v+m}$, stays in compartment D for an average time of $\frac{1}{w\eta+m}$, and causes α new infections per time unit. After moving to compartment R, he will neither infect any susceptible population, nor get reinfected because all other population are susceptible. Thus, total number of new infections that one infected individual causes in a susceptible population during his life time is:

$$\alpha \cdot \frac{1}{v+m} + \frac{v}{v+m}\left(\alpha \cdot \frac{1}{w\eta+m}\right) \equiv \mathscr{R}_c \tag{4.9}$$

which is exactly the value of $\mathscr{R}_c$ in (4.8).

Next, we show that the reproduction number $\mathscr{R}_c$ is a threshold quantity for asymptotic stability properties at the equilibrium points of the epidemic population.

Proposition 4.3 *The HCV transmission model* (4.1)-(4.4) *has a unique disease-free equilibrium (DFE)* $\mathcal{E} = \left(\frac{\Lambda}{\mu}, 0, 0, 0\right)$. *The disease-free equilibrium* $\mathcal{E}$ *is locally and globally asymptotically stable if* $\mathscr{R}_c < 1$.

Proposition 4.4 *There exists a unique endemic equilibrium (EE) if* $\mathscr{R}_c > 1$, *and no EE exists if* $\mathscr{R}_c < 1$.

Propositions 4.3 and 4.4 imply that $\mathscr{R}_c = 1$ is the critical value: the epidemic will eventually be eliminated only when $\mathscr{R}_c < 1$ with sufficient intervention efforts;

if $\mathscr{R}_c > 1$, the epidemic could converge to a positive endemic equilibrium. From the calculation of reproduction number $\mathscr{R}_c$, we observe that neither treatment nor screening alone can achieve the disease elimination in long run. This is because in the absence of either screening or treatment:

$$\mathscr{R}_c(v, w)\big|_{v=0,\text{or } w=0} = \frac{\alpha}{m} > 1,$$

as $\alpha > m$ (Assumption 4.1). By letting $\mathscr{R}_c < 1$, we obtain simple lower bounds for constant screening rate $v > \alpha - m$ and treatment rate $w > \frac{\alpha - m}{\eta}$ to achieve disease elimination in long term.

From Propositions 4.3 and 4.4, we know that the system equilibrium changes with the value of reproduction number $\mathscr{R}_c$, and such a phenomenon is called *bifurcation* (117). Moreover, with the base case parameter assumption $\alpha \geq \beta$ (Assumption 4.1), the bifurcation is *forward* at the *bifurcation point* $\mathscr{R}_c = 1$; that is, the equilibrium changes if $\mathscr{R}_c$ increases from 1. Of theoretical interest, in the following corollary, we remark that a *backward bifurcation* is possible if reinfection rate is higher than infection rate, i.e., the equilibrium changes if $\mathscr{R}$ decreases from 1, implying that there could exist at least one nontrivial endemic equilibria for $\mathscr{R}_c < 1$.

Corollary 4.1 *A necessary and sufficient condition for backward bifurcation is*

$$\beta > \alpha + \frac{(v+m)(w\eta+m)}{vw\eta}(\mu\mathscr{R}_c + \beta(1-\mathscr{R}_c)).$$

We remark that analyses of the critical value $\mathscr{R}_c$ and asymptotic stability properties provide qualitative insights and help us understand how quickly disease prevalence can be decreased and eliminated eventually in *infinite* time horizon. However, it has limited value of guiding specific actions to achieve certain hepatitis C elimination target while considering other important factors such as limited resources in finite time horizon. In the following sections, we turn to the optimal intervention problems that trade-off between the epidemic control and limited resources. We present the optimization problem formulations, discuss the optimal intervention policy structures,

showcase the numerical results, and discuss the policy implications in practice.

4.4 Optimal Resource Allocation: Final Prevalence Target Constraint

In the first global health sector strategy on viral hepatitis proposed by the WHO (169), the target is to eliminate viral hepatitis as a major public health problem by 2030. The strategy defines a set of quantitative targets for the next 15 years. In particular, the impact targets include 90% reduction in new incidences, and 65% reduction in liver-related mortality by 2030 compared with 2015; coverage targets include that 90% infected cases are diagnosed and that 80% of treatment-eligible patients are treated.

Motivated by the above targets stated in WHO strategy, we aim to solve the following optimization question: *In order to achieve a target disease prevalence (i.e., the number of total infected population), what is the optimal screening and treatment policy for 2015-2030 which has the minimum total cost*? Mathematically, we refer to the screening and treatment policy as the trajectories of $v(t)$ and $w(t)$ over time, respectively.

We let c_s^a denote the cost of HCV antibody test used for the screening in the general population. For those who show positive infection in the antibody test will undergo HCV RNA test for diagnosis at the testing cost of c_s^r. Let c_d represent the cost of each antiviral treatment. Then we have the following optimal control problem:

$$\text{(P-MinC)} \quad \min_{v,w} \quad \int_0^{t_f} \left[c_s^a \cdot (S+I+R) \cdot v + c_s^r \cdot I \cdot v + c_d \cdot D \cdot w\right] \, dt \tag{4.10}$$

$$s.t. \quad (4.1) - (4.6),$$

$$I(t_f) + D(t_f) \leq \rho_f, \tag{4.11}$$

$$v(t), w(t) \in [0,1], \ \forall t \in [0, t_f],$$

where ρ_f represents the target prevalence level to meet at the final time t_f.

Although the final prevalence is not explicitly stated as a target in the WHO strategy, we highlight that prevalence is indeed in line with the two impact target measures defined by the WHO—the numbers of liver-related deaths and new infections. In particular, given the setup of our HCV transmission model (4.1)-(4.4), the number of liver-related deaths is given by $(m-\mu)(I+D)$, which is simply proportional to prevalence $(I+D)$ by a constant factor. Moreover, the number of new infections $\varphi = \frac{(\alpha S+\beta R)(I+D)}{N}$ is highly sensitive to the prevalence $(I+D)$ when the $\frac{I+D}{N}$ is low, which can be seen from the following argument: we calculate the gradient $\frac{\partial \varphi}{\partial (I+D)} = \frac{(\alpha S+\beta R)(S+R)}{N^2} = \frac{(\alpha S+\beta R)}{N} \cdot \left(1-\frac{I+D}{N}\right)$, $\frac{\partial \varphi}{\partial S} = \frac{\alpha(R+(I+D))}{N} \cdot \frac{I+D}{N}$, and $\frac{\partial \varphi}{\partial R} = \frac{\alpha(S+(I+D))}{N} \cdot \frac{I+D}{N}$, and observe that $\frac{\partial \varphi}{\partial (I+D)}$ is approximately more than 10 times higher than $\left(\frac{\partial \varphi}{\partial S} + \frac{\partial \varphi}{\partial R}\right)$ when $\frac{I+D}{N} < 10\%$ as in most realistic HCV epidemic settings. Thus, we consider the final prevalence constraint in the optimal control problem (4.10) as a reasonable representation of the WHO targets. In the next result, we show the existence of optimal intervention policies to this formulation.

Proposition 4.5 *There exists an optimal control for problem P-MinC* (4.10)*.*

4.4.1 Structure of optimal control policies

In this section, we analytically examine the structure of the optimal intervention policies to the problem P-MinC (4.10) using the Pontryagin's Maximum Principle, which is a powerful mechanism used to establish the optimality conditions for optimal control problems (118, 190, 218).

For the convenience of following analyses, we formulate the problem in the form of Mayer-type problem; that is, we only minimize a terminal reward at the final time without the running cost during planning horizon $[0, t_f]$. This can be easily achieved with the expansion of state (S, I, D, R) by an additional state variable $Z(t)$, which captures the total cost at time t with the dynamics $\dot{Z} = c_s^a(I+D+R)v + c_s^a Iv + c_d Dw$ and the initial condition $Z(0) = 0$.

Then we have the following reformulation of (4.10) in the Mayer form with state $x(t) = (Z(t), S(t), I(t), D(t), R(t))$:

$$\begin{aligned} \text{(P-MinC-Mayer)} \quad \min_{(v,w)\in[0,1]^2} \quad & Z(t_f) && (4.12)\\ s.t. \quad & \dot{x} = f(x) + g_1(x)v + g_2(x)w && (4.13)\\ & x_0 = (0, S_0, I_0, D_0, R_0) \\ & I(t_f) + D(t_f) \leq \rho^{target} \end{aligned}$$

where

$$f(x) = \begin{bmatrix} 0 \\ \Lambda - \mu S - \frac{\alpha S(I+D)}{N} \\ \frac{(\alpha S+\beta R)(I+D)}{N} - mI \\ -mD \\ -\frac{\beta R(I+D)}{N} - \mu R \end{bmatrix}, \; g_1(x) = \begin{bmatrix} c_s^a(S+I+R) + c_s^r I \\ 0 \\ -I \\ I \\ 0 \end{bmatrix}, \; g_2(x) = \begin{bmatrix} c_d D \\ 0 \\ 0 \\ -\eta D \\ \eta D \end{bmatrix}. \tag{4.14}$$

Define the Hamiltonian $H = H(x, u, \lambda, t)$ as

$$H = \langle \lambda, (f(x) + g_1(x)v + g_2(x)w) \rangle, \tag{4.15}$$

where $\lambda = (\lambda_0, \lambda_1, \cdots, \lambda_4) \in \mathbb{R}^5$ are the *costates* (also referred to as "adjoint variables") associated with system dynamics (4.13).

First order necessary conditions of optimal control $u^*(t)$ are given by the Pontryagin's Maximum Principle (Theorem 2.2 and 3.3 in (190)) as follows. Let $(x^*(t), u^*(t))$ be an optimal pair, there exists a continuous and piecewise continuously differentiable vector function $\lambda(t) = (\lambda_0(t), \cdots, \lambda_4(t))$ satisfying the following:

(1) $\lambda(t) \not\equiv 0$,

(2) Costate equations:

$$\dot{\lambda} = -\frac{\partial H}{\partial x} = -\left\langle \lambda, \frac{\partial f(x^*)}{\partial x} + \frac{\partial g_1(x^*)}{\partial x} v^* + \frac{\partial g_2(x^*)}{\partial x} w^* \right\rangle, \tag{4.16}$$

with transversality conditions:

$$\begin{aligned}&\lambda_0(t_f) = 1, \lambda_1(t_f) = \lambda_4(t_f) = 0,\\ &\lambda_2(t_f) = \lambda_3(t_f) = k, \text{ for some } k \geq 0 \text{ and } k = 0 \text{ if } I^*(t_f) + D^*(t_f) < \rho^{target},\end{aligned} \tag{4.17}$$

(3) $u^*(t)$ minimizes $H(x^*(t), u, \lambda(t), t)$ for any time t over the compact interval $U = [0, 1] \times [0, 1]$. That is,

$$H(x^*(t), u^*(t), \lambda(t), t) \leq H(x^*(t), u(t), \lambda(t), t) \text{ for all } u \in U.$$

Note that the Hamiltonian H is linear in control $u = (v, w)$ in our formulation, which implies that minimum H can be attained at the boundary values of control u depending on the sign of gradient of $\frac{\partial H}{\partial u}$. In particular, the optimal control $u^*(t)$ is characterized in the following proposition.

Proposition 4.6 *Let $(x^*(t), u^*(t))$ be an optimal pair. Costates $\lambda(t)$ is determined by* (4.16)-(4.17)*. Then, the optimal control $u^*(t) = (v^*(t), w^*(t))$ is determined as follows:*

$$v^*(t) = \begin{cases} 0 & \text{if } \Phi_1(t) > 0 \\ \text{undetermined} & \text{if } \Phi_1(t) = 0 \\ 1 & \text{if } \Phi_1(t) < 0 \end{cases}, \quad w^*(t) = \begin{cases} 0 & \text{if } \Phi_2(t) > 0 \\ \text{undetermined} & \text{if } \Phi_2(t) = 0 \\ 1 & \text{if } \Phi_2(t) < 0 \end{cases} \tag{4.18}$$

where switching functions $\Phi_1(t)$ *and* $\Phi_2(t)$ *are defined as*

$$\begin{aligned}\Phi_1(t) =& c_s^r(S^* + I^* + R^*) + c_s^a I^* + (\lambda_3 - \lambda_2)I^*,\\ \Phi_2(t) =& c_d D^* + (\lambda_4 - \lambda_3)\eta D^*.\end{aligned}$$

Conditions in (4.18) shed some lights on the basic policy structure. Most of the time, optimal control is at the boundary value, and can switch to the opposite

boundary at some critical time τ, for example, from 1 to 0 when the switching function $\Phi_i(\tau)$ across the value 0 from negative to positive. Controls switching between the minimum and maximum values are called the *bang-bang policy*.

However, the optimal controls are not directly determined by condition (4.18) if the switching functions vanish over an open interval, i.e., $\Phi_i(\tau) = 0$, $\tau \in [\tau_1, \tau_2]$, for some $\tau_2 > \tau_1$. In this case, on $[\tau_1, \tau_2]$, $\frac{\partial^2 H}{\partial u^2} \equiv 0$ which is singular and renders the first-order optimality condition (i.e., the maximum principle) informative. As such, the problem is called *singular*. To determine the singular controls, additional necessary conditions are needed to fully characterize the optimal control. For example, one can utilize the facts $\dot{\Phi}_i \equiv 0$, $\ddot{\Phi}_i \equiv 0$, $\ldots, \Phi_i^{(2k)} \equiv 0$ for some $k \geq 1$, and check the second-order optimality condition, $(-1)^k \frac{d}{du}\left[\left(\frac{d}{dt}\right)^{2k} \frac{\partial H}{\partial u}\right] \geq 0$, known as the *generalized Legendre-Clebsch condition* (34).

On the other hand, conditions in (4.18) only indicate that there *may* exist singular arcs, but do not guarantee the existence of singular arcs. The existence of singular solutions could also depend on specific parameter values or the initial condition of the problem (c.f., (158)).

We remark that if the cost function has quadratic terms with respect to control variable u such as $((S + I + R) \cdot v)^2$ and $(D \cdot w)^2$, the singularity issues can be completely resolved. However, it could be more difficult to interpret a quadratic monetary value in our problem context, and could be also more difficult to obtain practically meaningful estimates of the coefficients for the quadratic terms. We shall discuss more on the forms of objective functions in our numerical result section.

To summarize, because the cost function and the system dynamics are both linear in control u in the problem P-MinCost (4.12), the optimal control admits a *generalized bang-bang policy*, possibly in combination with singular arcs.

In the following proposition, we explore the possibility of singular controls in two special cases of our problem.

Proposition 4.7 *Control v and w cannot be singular simultaneously in the following cases:*

(1) In a simplified model without the Recovered compartment. That is, no reinfection is considered.

(2) In the full model (4.12) *when $\alpha = \beta$. That is, reinfection rate is as high as the infection rates for the susceptible population.*

Analytical solutions to singular controls are in general challenging and intractable. By Proposition 4.7, we can only rule out the simultaneous singularity of optimal screening and treatment policies in two cases with either negligible reinfections or very high reinfection rates; singular arcs for either intervention may still exist in the original problem. In fact, in our computational experiments, we fail to solve the problem (4.10) using the numerical solver, which fundamentally relies on the first-order optimality conditions for general non-linear optimization problems. Thus, we conjecture that singular arcs exist for the problem (4.10), which also renders numerical solutions intractable.

4.4.2 Fixed (and constant) screening rate

Instead of investigating the exact singular solutions of theoretical interest which could be highly dynamic and complex, in this section, we focus on a class of policies with a simpler structure, which can be more practical for implementation, and may also possibly circumvent the singularity issues.

To simplify the optimal control problem, we keep only one type of intervention as the control variable in the model, and set the other at a given value. Specifically, we choose treatment as the control variable to be optimized and set screening at a fixed rate. Considering that the target population of screening has a much larger scale than that of treatment, we believe it is more practical to first fix the screening policy

with a simple and practical structure (e.g., constant rate, or piece-wise constant rate for several phases), and then determine the corresponding optimal treatment strategy possibly with a more flexible structure focusing on only a smaller patient group.

For a given screening rate $v^\circ(t)$, problem (4.12) is reduced to the following. For state $x = (Z, S, I, D, R)$,

$$\min_{w(t)\in[0,1]} \quad Z(t_f) \tag{4.19}$$

$$s.t. \quad \dot{x} = f(x) + g(x) \cdot w \tag{4.20}$$

$$x_0 = (0, S_0, I_0, D_0, R_0)$$

$$I(t_f) + D(t_f) \leq \rho^{target}$$

where

$$f(x) = \begin{bmatrix} c_s^a(I+S+R)v^\circ + c_s^r I v^\circ \\ \Lambda - \mu S - \alpha S \frac{I+D}{N} \\ (\alpha S + \beta R)\frac{I+D}{N} - (m+v^\circ)I \\ v^\circ I - mD \\ -\mu R - \beta R \frac{I+D}{N} \end{bmatrix}, \; g(x) = \begin{bmatrix} c_d D \\ 0 \\ 0 \\ -\eta D \\ \eta D \end{bmatrix} \tag{4.21}$$

In Proposition 4.8, we show that fixing screening rate at a given value $v^\circ(t)$, not necessarily a constant value, indeed excludes the existence of the singular arcs for the treatment policy. That is, for a given screening policy, the optimal treatment policy is always a *pure bang-bang* policy, meaning either no treatment at all (at the minimum rate $w(t) = 0$) or treating all existing diagnosed patients (at the maximum rate $w(t) = 1$). This simple policy structure is appealing form an implementation perspective as well, as the decision now becomes when to start or stop treatment for all diagnosed patients.

Proposition 4.8 *For any given screening rate $v^\circ(t) \not\equiv 0$ for all time t, the optimal treatment rate $u^*(t)$ is a pure bang-bang policy without singular arcs.*

Next, we consider a simple screening policy with a constant screening rate $v^\circ(t) \equiv v^c$. Recall that the optimal control problem is subject to the target constraint of the final prevalence. Clearly, all constant screening rates v^c are not admissible to the problem (4.19). To characterize the feasible range of v^c, we first show that, given the constant screening rate, the minimum final prevalence one can achieve is nonincreasing in v^c in Proposition 4.9. Then it follows that there exists a *minimal screening rate* $\underline{v}^c$ such that the final prevalence constraint (4.11) is satisfied (Corollary 4.2). Moreover, we can show that, in Proposition 4.10, more treatment efforts lead to lower prevalence, with under certain simplification conditions. As a result, a full treatment policy, i.e., $w(t) = 1$ for all time t, becomes the only feasible and thus the optimal treatment policy given the minimal screening rate $\underline{v}^c$ to achieve the prevalence target (Corollary 4.3). Indeed, as shown in our numerical results, the intervention policy with screening $v(t) \equiv \underline{v}^c$ and treatment $w(t) \equiv 1$ is optimal with the lowest cost among all policies with a constant screening rate.

Proposition 4.9 *Consider the following optimal control problem to minimize final prevalence under the given constant screening rate* $v^c \in [0, 1]$*:*

$$J^*(v^c) = \min_{w(t)\in[0,1]} \{J(v^c) := I(t_f) + D(t_f) : (4.1) - (4.6)\}. \tag{4.22}$$

The optimal terminal prevalence $J^*(v^c)$ *is nonincreasing in screening rate* v^c*.*

Corollary 4.2 *Consider problem* (4.19) *with a constant screening rate, i.e.,* $v^\circ(t) \equiv v^c$ *for all time* t *for some* $v^c > 0$*. There exists a minimum constant screening rate* $\underline{v}^c$ *to achieve the terminal prevalence target. That is, problem* (4.19) *with constant screening rate* v^c *is feasible only for* $v^c \in [\underline{v}^c, 1]$*.*

Proposition 4.10 *Consider a stationary population by assuming* $\Lambda = \mu N_0$ *and* $m = \mu$*. Let two state trajectories* $x_i(t) = (S_i(t), I_i(t), D_i(t), R_i(t))$ *determined by the control* $u_i(t) = (v(t), w_i(t))$ *for* $i = 1, 2$*. If* $w_1(t)D_1(t) \le w_2(t)D_2(t)$*, then* $S_1(t) \le S_2(t)$*,* $R_1(t) \le R_2(t)$*, and* $I_1(t) + D_1(t) \ge I_2(t) + D_2(t)$*.*

Corollary 4.3 *Given the minimal constant screening rate* $\underline{v}^c$*, treatment rate* $w^*(t) \equiv 1$ *is the optimal and only feasible treatment policy to achieve the final prevalence target.*

4.4.3 Parameter estimation

In our numerical analyses, we consider the HCV epidemic in India population. HCV elimination in India is an interesting study case because the disease burden in India is high (with over 6 million people infected with chronic HCV), and effective oral therapies are available at reduced costs (216). India government has started working with the WHO to develop intervention plans for HCV elimination. Motivated by such a realistic policy-making problem, we parameterize our model base on the India setting (e.g., epidemic population profile, intervention costs), numerically solve the optimal control problem, and obtain managerial insights on the intervention strategies from the optimal policies.

Population profile. Since the model is scalable with the compartment sizes, we run the results for N_0 =1 million initial population. Although the overall nation-wide prevalence rate is low (0.68%) as diluted by the large Indian population, the epidemic is very severe in certain regions, like Punjab with a high prevalence rate of 5.2% (206). Our analyses will focus on such targeted area and population. Awareness of infection is low in India; it is estimated that only 5% of infected people are diagnosed (228). As we do not differentiate the compartments by age groups, we use the crude death rate estimate for overall Indian population as the background mortality μ (214). To estimate an aggregated excess mortality due to HCV infection (i.e., for I and D compartments), we use the weighted average of liver-related mortality by liver disease stage with respect to their distribution in India population. We assume a constant birth rate $\Lambda = \mu N_0$.

Screening and treatment. HCV screening includes anti-HCV antibody test at a cost of $5, and RNA test at a cost of $50 in India (as per communication with

practitioners in India). The cost of DAA treatment in India has now reduced to $300 for 12-week treatment (as per communication with WHO-India). The new DAA has shown very high efficacy with >95% SVR rate for all fibrosis stages and genotypes (238). Considering the sub-optimal uses and adherence to treatment in actual clinical practice, we use $\eta = 0.9$ as a conservative SVR estimate in our model.

Table 4.2: Parameter values.

Parameter	Variable	Value	Reference
Total population	N_0	1,000,000	
Initial prevalence		5.2%	(206)
Initial awareness %		5%	(99)
Current intervention			
Annual screening rate		0.6%	(99)
Annual treatment rate		0.2%	(99)
Mortality (annual)			
Background mortality	μ	0.73%	(214)
Excess mortality	$m - \mu$	0.8%	weighted average of liver-related mortality by liver disease stages[1]
Screening cost			
Antibody test	c_s^a	$5	per communication with clinicians in India
RNA test	c_s^r	$50	per communication with clinicians in India
DAA treatment cost	c_d	$300	per communication with WHO-India
SVR rate	η	0.9	(238)
Infection rate	α	0.0167	Estimated through calibration
Relative susceptibility of treated patients	κ_r	30%	(100, 89)

[1] No liver-related mortality for F0-F3 fibrosis stages, annual mortality 0.051 for F4 (6.89% of the total hepatitis C-infected population), 0.265 for decompensated cirrhosis (1.43%) (54), and 0.427 for liver cancer (0.18%) (100).

Calibration of infection rate. As the direct estimate of infection rates α and β are not available from the literature, we use Martin's approach (157) to calibrate the parameters as follows. We assume a $70\%(= 1 - \kappa_r)$ reduction of reinfection rate for those who have been successfully treated compared with the new infections (i.e., $\beta = \kappa_r \alpha$) as in (100), in line with the findings from clinical studies that treated patients are about 4 times less likely to develop infection than those infected for the first time (89). Using the number of newly diagnosed infections and treatment estimated in (99), we calculate the status quo annual screening rate to be 0.6% and annual treatment rate 0.2%. Then we run the transmission model with the status quo

intervention levels for long enough until the equilibrium state is reached for any given infection rate α value, and calibrate the α value such that the equilibrium prevalence rate under the status quo interventions is inline with the current prevalence estimate (5.2%) in India.

Figure 4.2 shows the relation between the equilibrium prevalence and infection rate α. In line with our results in Propositions 4.3 and 4.4, when α is smaller than the critical value $\alpha^0 = 0.0158$ so that the reproduction number $\mathscr{R}_c < 1$, the dynamic system will eventually converge to a disease-free equilibrium. An endemic equilibrium bifurcates forwardly at $\alpha = \alpha^0$ and $\mathscr{R}_c = 1$. That is, the endemic equilibrium prevalence will increase with α value when $\alpha > \alpha^0$ and $\mathscr{R}_c > 1$.

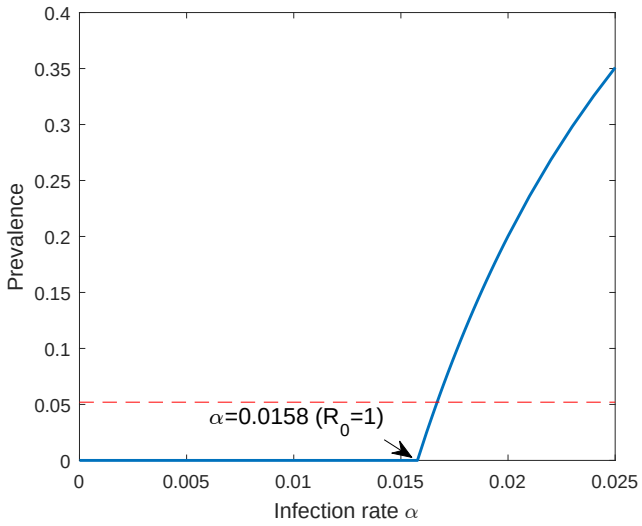


Figure 4.2: Calibration of infection rate α.

Note. The red dashed-line represents the calibration target prevalence. The intersection point with the solid curve represents the base case value of infection rate α used in our numerical analyses.

We remark that the infection rate α is a sensitive parameter affecting the equilibrium prevalence, disease transmission dynamics, and corresponding intervention policies. Our estimate of α from the calibration serves as an initial basis for the following analyses of the optimal intervention policy in HCV elimination; a more

accurate estimate of α is possible if more data on the epidemic (e.g., longitudinal incidence and prevalence statistics) can also be utilized in the calibration. We will examine the effect of variability of α values on the policy structure in more details in our numerical results.

4.4.4 Numerical results

The WHO 2016-2021 global health sector strategy on viral hepatitis (215) presents a set of targets to achieve by 2030, including 65% reduction of liver related mortality and 90% reduction of new incidences. As the number of the infected patients is the major quantity driving both mortality and new incidence (see discussion in Section 4.4), in our base case results we set the target of reducing the number of infected patients by 65% at 2030 compared with the initial number of the infected at 2015 as the final state constraint (4.11) in the model.

Optimal screening and treatment policies are numerically solved using the MATLAB package GPOPS-II (172). GPOPS-II employs the *direct method* for solving optimal control problems, i.e., discretize-then-optimize, in contrast to the *indirect method* which first derives the optimality conditions and then iterates the solutions in discrete time until the optimality conditions are satisfied, i.e., optimize-then-discretize, such as the multiple-shooting method (178). In particular, it approximates the continuous-time optimal control as sparse non-linear programming (NLP) using different collocation schemes, and then solves the NLP with NLP solvers. Computation time for all problem instances are less than one minute.

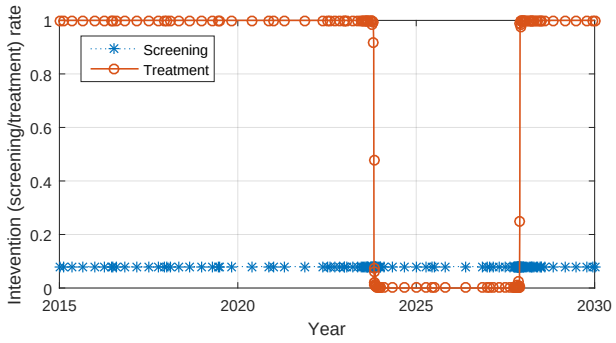
4.4.4.1 Fixed constant screening rate

As the base case result, we choose a constant annual screening rate $v^c = 8\%$ throughout 2015-2030 and solve the optimal treatment policy accordingly. That is, screening covers 8% of all population every year except those who have been diagnosed with HCV infection (i.e., compartment D). This particular choice of screening rate is made

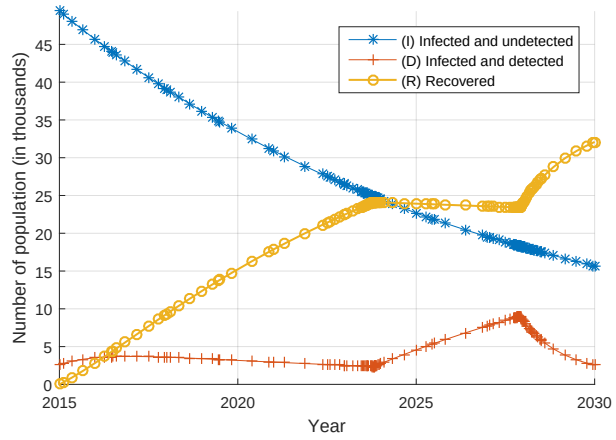
for the purpose of exploring and presenting the underlying rationales of the optimal treatment policy. Figure 4.3 presents the optimal treatment policy and the corresponding state trajectory over time. The markers shown in the figure represent the discretization points for time discretization which are adaptively determined by the solver (e.g., smaller discretization in the region with more rapid changes of state or control values). We observe that optimal treatment is indeed a pure bang-bang policy without singular arcs, which numerically verifies our result in Proposition 4.8.

Figure 4.3 shows that given the constant screening rate, the number of the undiagnosed infections (compartment I) is decreasing steadily. The treatment rate affects the number of diagnosed infections and those recovered differently in three time phases: (1) before the year 2024, treatment at the maximum rate 100% strives to move the diagnosed infections to the recovered state (i.e., from compartment D to R), and maintain the number of diagnosed infections under control; (2) during the years 2024-2028, treatment pauses, and the number of diagnosed infections increases as more infections continue to be diagnosed but no treatment is available during this period; and (3) after the year 2028, treatment resumes at the maximum rate 100% and successfully reduces the prevalence to the targeted level at the end of 2030.

Interestingly, the optimal treatment policy exhibits an *on-off-on* structure to minimize the total screening and treatment cost. To understand the underlying reason of such a policy structure, we first highlight the multiple effects of treatment. On the one hand, treatment immediately reduces the disease prevalence, and the reduced prevalence in turn results in fewer new infections in long term; on the other hand, since the treated and recovered people also constitute a part of screening target population, the growing number of recovered people could increase the total screening cost. Then, the optimal treatment policy (Figure 4.3A) can be interpreted as follows. The early treatment in the first "on"-segment can be considered as the "treatment as prevention". That is, it helps reduce future new infections by reducing prevalence



(A) Screening rate and the optimal treatment rate.



(B) State trajectory.

Figure 4.3: Optimal treatment policy and the state trajectory at a fixed annual screening rate of 8%.

in early periods and leads to potential savings of treatment cost in future. When such preventive benefits are offset by the increasing screening cost as the screening target population grows, treatment becomes less attractive, or even stops. Delaying treatment is feasible only if it is still possible to meet the final prevalence target by implementing the maximum treatment rate—the second "on"-segment—later and throughout the remaining time. As such, the three stages of the on-off-on structure can be interpreted as the *preventive treatment*, *feasible delay*, and *target-approaching treatment* stages, respectively.

However, it is not guaranteed that the optimal treatment policy of the bang-bang

policy type always consists of such three stages. For different model parameter settings, other switching structures, such as *on-off* or *off-on* structures, are also feasible. In the following, we perform a series of sensitivity analyses and discuss the effects of several major model parameters on the switching structure of the optimal treatment policy.

- **Effect of infection rate** α (compared to mortality m). From earlier discussion on reproduction number $\mathscr{R}_c$, in the absence of treatment, we know that $\mathscr{R}_c$ reduces to $\frac{\alpha}{m}$. A lower infection rate α results in a smaller reproduction number $\mathscr{R}_c$ which is closer to 1, implying a slow spread of the disease; in this case, treatment may not exhibit a strong preventive value on the epidemic control, which indicates that early treatment is not necessarily optimal. As shown in Figure 4.4, when infection rate α is low (the upper panel), optimal treatment policy is to delay the treatment until necessary for reaching the final prevalence target; when infection rate α is high with a faster disease transmission (the lower panel), upfront and non-stop treatment could be more efficient to prevent the new infections, and also more economical to achieve the final prevalence target.

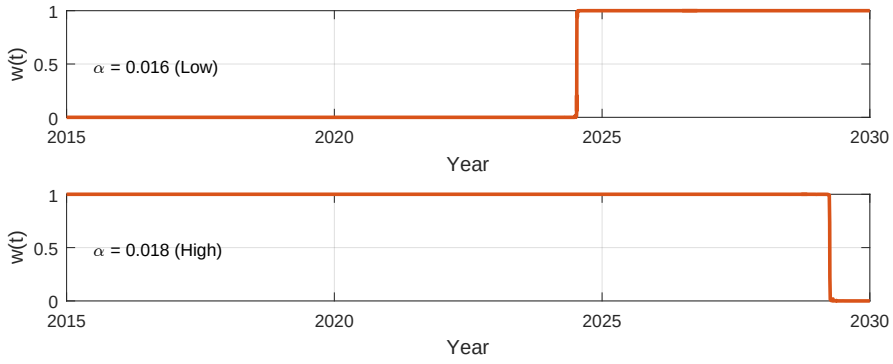


Figure 4.4: Effect of infection rate α on the optimal treatment policy.

- **Effect of antibody testing cost** c_s^a for general screening target population.

Figure 4.5 shows the optimal treatment policies under various values of antibody testing cost c_s^a. It presents a clear pattern: a higher antibody testing cost favors later rather than early treatment. With a higher antibody testing cost, early treatment is not favored because it increases the screening target population in early periods, which will contribute to higher screening cost throughout the horizon. Then the potential savings of the preventive treatment is counterbalanced by such an increase of total screening cost. Therefore, it would be better to maintain less early treatment as the antibody testing cost increases.

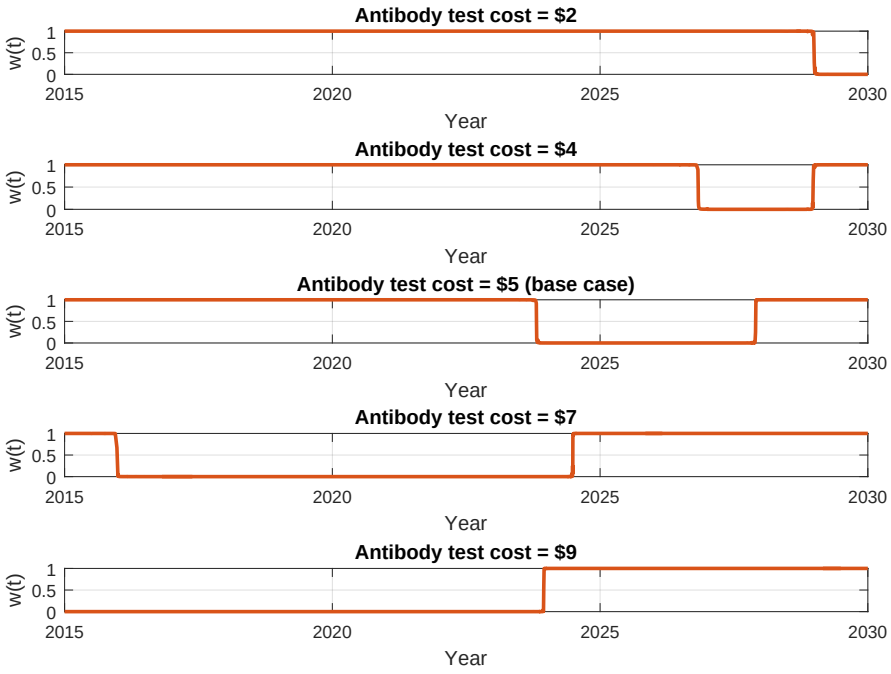


Figure 4.5: Effect of antibody testing cost c_s^a on the optimal treatment policy.

- **Effect of RNA testing cost** c_s^r for detected infections. RNA testing cost c_s^r shows an opposite effect on results compared to the antibody testing cost c_s^a in Figure 4.6. Unlike the HCV antibody testing for general population, RNA tests are only performed on those who have a positive antibody test. In other words, RNA test costs are only attributed to the detected $v(t) \cdot I(t)$ cases, and

can be considered as an additional penalty cost to the number of the infected $I(t)$ in the cost function. To achieve a lower value of $I(t)$, i.e., a smaller size of compartment I, it is critical to reduce the incoming flow—new infections or reinfections—to the compartment I, which indeed strengthens the preventive value of early treatment. Therefore, more early treatment would be expected when RNA testing cost becomes higher.

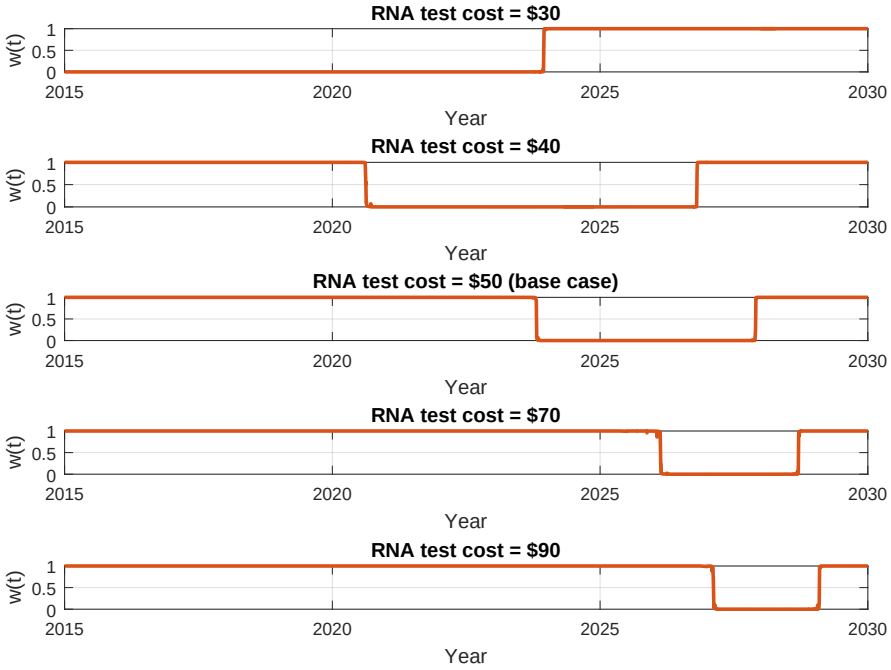


Figure 4.6: Effect of RNA testing cost c_s^r on the optimal treatment policy.

- **Discounted cost.** Many economic analyses consider time discounting of future cost. To account for time discounting in the optimal policy, the only change needed is to include a discounting factor in the objective function of our model as follows:

$$\min \int_0^{t_f} \frac{1}{(1+r)^t} \left[c_s^a(S(t)+I(t)+R(t))v^\circ + c_s^r I(t)v^\circ + c_d D(t)w(t)\right] dt,$$

where r represents the annual discount rate. Since lower weights are now placed

on the costs that are incurred in future, treatment efforts are shifting towards later periods (Figure 4.7). For a small discount rate of 0.03% per year, early treatment is still needed in the optimal policy; whereas for a commonly used discount rate of 3% per year, early treatment is no longer desirable and all treatment efforts shift towards later periods.

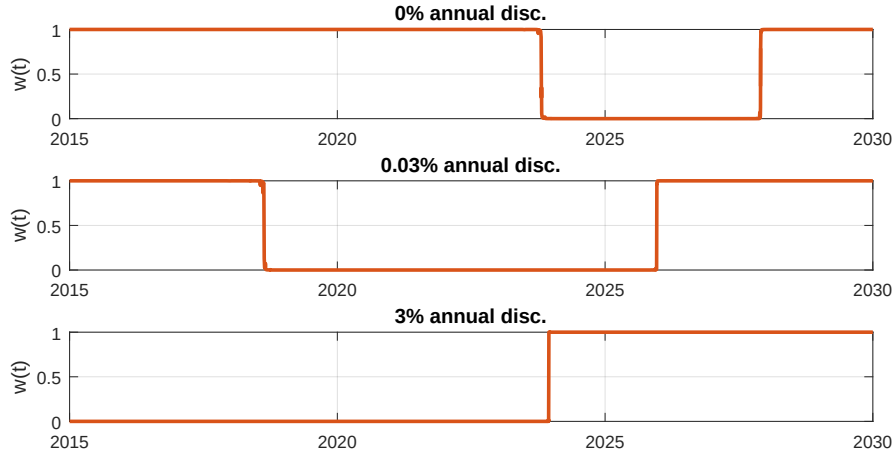
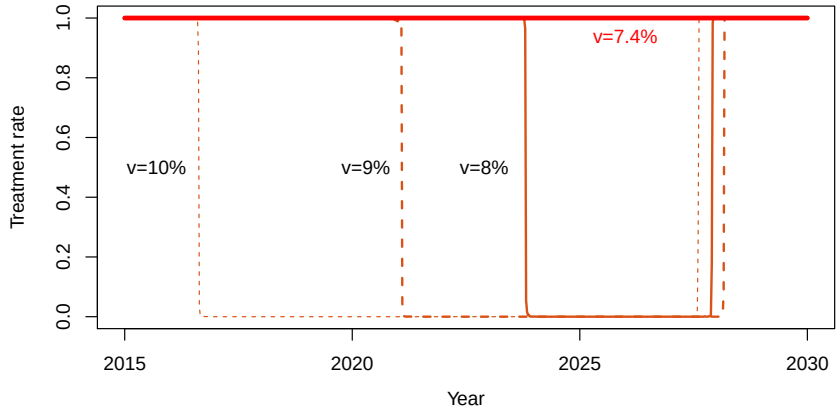


Figure 4.7: Effect of time discounting factor on the treatment policy.

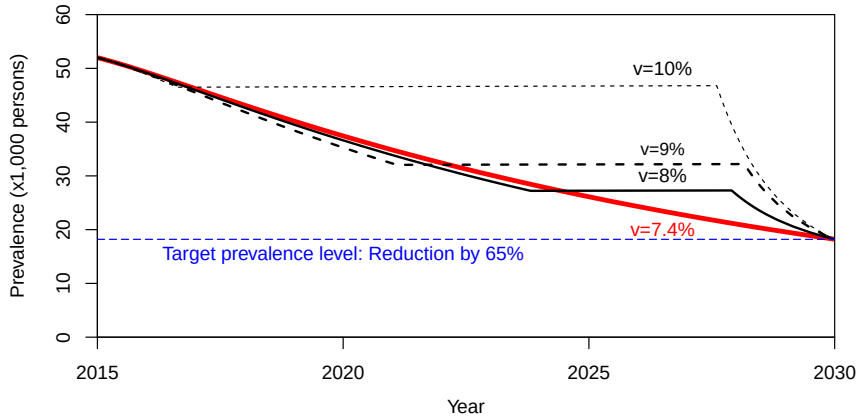
4.4.4.2 Minimal screening rate

As we discussed in Section 4.4.4.1, the *feasible delay* stage is possible only if implementing the maximum treatment rate throughout the remaining time can bring the prevalence down to the target value. In other words, the feasibility question reduces to whether the screening has detected sufficiently many patients for treatment. If the screening rate is too low, many infected people still remain undiagnosed and do not have access to treatment. In such a case, treatment at the maximum rate cannot reduce the total prevalence to the target level. On the other hand, if screening at a higher rate is able to detect more infected patients than needed, the optimal treatment will only cover the exact amount that is needed to reduce the prevalence to the target level, resulting in the delay periods of treatment as we observed earlier.

As shown in Figure 4.8A, the time window of treatment delay shortens as the



(A) Optimal treatment rate



(B) Prevalence

Figure 4.8: Optimal treatment policy and prevalence for different constant screening rates.

screening rate decreases. When the screening rate decreases to a critical value $\underline{v}^c$, *the minimal screening rate*, there is no delaying periods and the optimal treatment policy is to simply treat every diagnosed patients with the maximum rate $w^{max} = 100\%$ all the time. As such, the prevalence continues decreasing from the beginning until reaching to the final target prevalence at the end of the horizon (Figure 4.8B).

More interestingly, we find that the minimal screening rate $\underline{v}^c$ with the maximum treatment rate w^{max} indeed results in the lowest total cost. Figure 4.9 shows that the total cost increases with the screening rate, for different prevalence reduction targets.

The dashed line indicates that the target prevalence cannot be achieved given that screening rate value. These results imply that any constant screening rate higher than $\underline{v}^c$ (corresponding to the leftmost point of each solid curve) is deemed *unnecessary*. The high screening rate could lead to more diagnosed cases than needed to achieve the prevalence target. That is, the extra spending on screening has no contribution in achieving the target prevalence. Thus, maintaining screening rate at the minimum but necessary level is deemed the most efficient way of utilizing resources to achieve the prevalence reduction target.

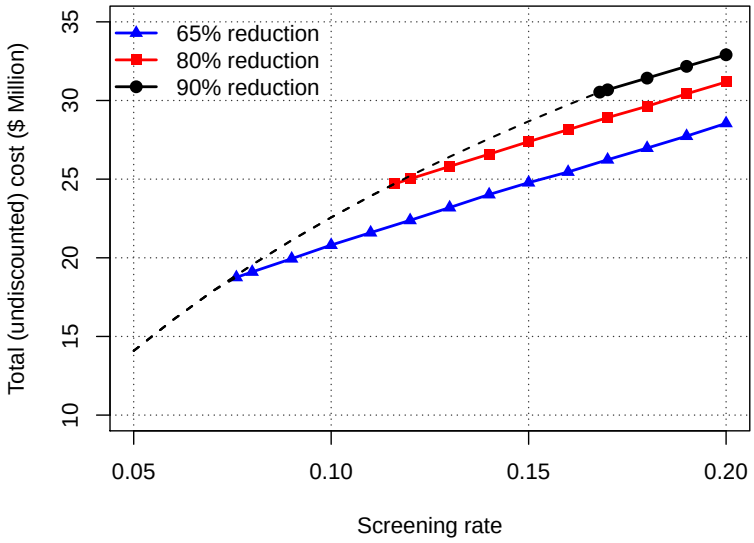


Figure 4.9: Total cost for screening and treatment by constant screening rate for different prevalence reduction targets.

Comparisons of results across different prevalence reduction targets (in Figure 4.9) also highlight the importance of screening on the HCV elimination. To achieve 90% reduction of prevalence, the minimal screening rate needs to be more than twice as high as that for 65% reduction. The minimal screening rate is not simply linearly scaled with the target prevalence reduction, because when the prevalence becomes

lower, it is more difficult to detect new infected cases unless screening more aggressively. From the comparisons, we also find that the total cost increment is largely attributed to the increased screening rate.

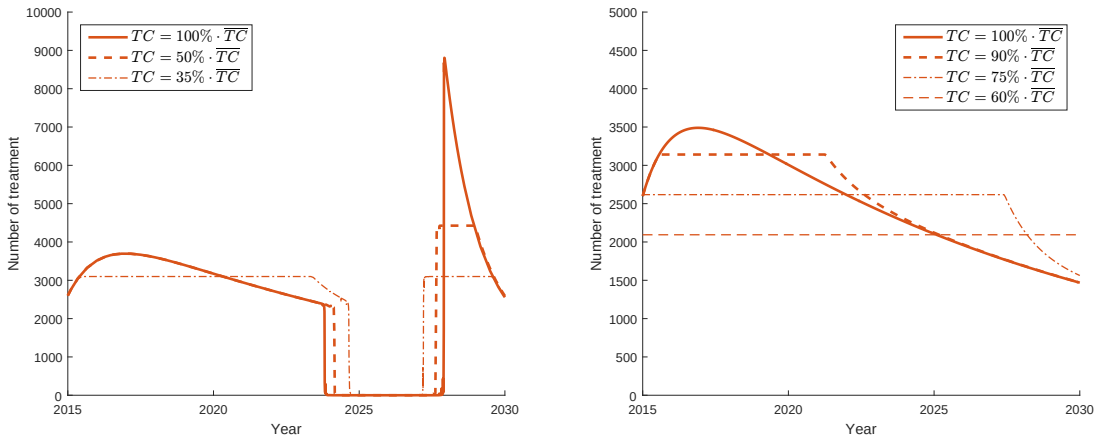
4.4.4.3 Treatment capacity

While treatment rate remains constant at the maximum rate of 100% whenever the treatment is not delayed, obviously the actual number of people receiving treatment is not stationary over time due to the dynamic size of the diagnosed population in compartment D. In practice, treatment capacity may not always meet the surge of high treatment volume. To capture such a constraint, we include the following constraint to the problem formulation P-MinC (4.10):

$$w(t)D(t) \leq TC, \ \forall t, \tag{4.23}$$

where TC represents the treatment capacity as a given parameter.

For a given constant screening rate v^c, we solve the problem with treatment capacity $TC = \gamma \cdot \overline{TC}$ with varies value of $\gamma < 1$ (e.g., $\gamma = 90\%, 80\%$) and $\overline{TC}$ defined as the maximum value of $D^*(t)$ when the problem is solved without treatment capacity constraint (4.23).



(A) Fixed $v^c = 8\%$ (B) $v^c = \underline{v}^c$ for 65% prevalence reduction

Figure 4.10: Number of treatment under various treatment capacities.

When screening rate is above the minimal rate (for the problem with unlimited treatment), i.e., $v^c > \underline{v}^c$, a large amount of patients are treated at the beginning of the target-approaching treatment stage (Figure 4.10A). Such a big surge of the treatment volume is smoothed out by the treatment capacity constraint. With limited treatment capacity, treatment is extended for longer periods such that the final prevalence target is still achievable.

When the minimal screening rate $\underline{v}^c$ is chosen, the additional treatment capacity constraint could render the problem infeasible, because $w^*(t) = w^{max}$ is the only feasible treatment policy in this case. Thus it would be interesting to investigate the trade-off between the practical treatment capacity constraint and the final prevalence target constraint. We evaluate the treatment policies $w(t) = \min\{\frac{TC}{D(t)}, 1\}$ with various treatment capacity TC (illustrated in Figure 4.10B). Table 4.3 shows that reducing treatment capacity to $75\% \cdot \overline{TC}$ results in a less than 0.5% gap from the 65% prevalence reduction target. For a higher reduction target of 90%, the final prevalence is more robust against the treatment capacity: a similar feasibility gap can be achieved by a treatment capacity $65\% \cdot \overline{TC}$.

4.4.4.4 Alternative formulations of cost functions

Quadratic cost structure. In many economic and engineering applications, quadratic cost structure is a popular choice for the objective cost function of optimal control models (118). Quadratic cost function has two appealing features. From an empirical perspective, quadratic cost with respect to control variables represents a *scaling-up cost* for increasing the control efforts. It makes sense in the context of disease screening: cost for screening 10% of the population is likely to be higher than twice of the cost for screening 5% of the population, as increasing the coverage would require extra manpower and resources in addition to the cost of tests themselves.

From the mathematical perspective, quadratic cost function is smooth and convex

Table 4.3: Effect of treatment capacity on the final prevalence reduction.

Prevalence reduction target	Treatment capacity (γ)	Max treatment number ($\gamma \cdot \overline{TC}$)	Prevalence reduction (compared to initial prevalence)	Total cost ($ Million)
65% reduction[1]	100%	3,489	65.0%	18.6
	95%	3,316	65.0%	18.6
	90%	3,141	65.0%	18.6
	85%	2,967	64.9%	18.6
	80%	2,792	64.7%	18.5
	75%	2,618	64.4%	18.5
	70%	2,443	62.8%	18.2
	65%	2,269	58.3%	17.4
	60%	2,094	53.7%	16.6
	55%	1,920	49.2%	15.8
	50%	1,745	44.6%	15.0
90% reduction[2]	100%	6,613	90.0%	30.4
	95%	6,282	90.0%	30.4
	90%	5,952	90.0%	30.4
	85%	5,621	90.0%	30.4
	80%	5,290	90.0%	30.4
	75%	4,960	89.9%	30.4
	70%	4,629	89.8%	30.4
	65%	4,298	89.6%	30.4
	60%	3,968	89.3%	30.3
	55%	3,637	88.5%	30.2
	50%	3,307	85.1%	29.6

[1] At minimal constant screening rate $\underline{v}^c = 7.4\%$.
[2] At minimal constant screening rate $\underline{v}^c = 16.7\%$.

which has many desirable properties. In particular, in our optimal control problem, a quadratic cost function will rule out the singularity issue, which makes it possible to directly compute the optimal screening and treatment policy at the same time. However, one challenge in implementing the quadratic cost function in our problem is the lack of realistic estimates for the coefficient of the quadratic terms.

In the following, we present the optimal policy with quadratic cost function based on assumed values for the scaling-up costs, and discuss the policy insights from the

results. In particular, we consider the following objective function:

$$\min \int_0^{t_f} \left\{ q_s[(S+I+R)v]^2 + [c_s^a(S+I+R)v + c_s^r Iv] + q_d(Dw)^2 + c_d Dw \right\} dt, \tag{4.24}$$

where q_s and q_d represent the scaling-up costs (\$ per person-squared) for screening and treatment, respectively. To determine scaling-up costs for numerical analysis, we calculate the base value of q_s such that, with the initial population size, the quadratic cost $q_s[(S+I+R)v]^2$ is comparable to and has a similar order of magnitude as the linear cost term $c_s^a(S+I+R)v + c_s^r Iv$. The same method is applied for calculating the treatment scaling-up cost q_d. In the following numerical results, we use $q_s^0 = 5 \times 10^{-5}$ and $q_d^0 = 5 \times 10^{-2}$ as the baseline values, and vary the value by an order of 100 to obtain the low and high values, respectively, i.e., $\frac{q_s^0}{100}$ and $\frac{q_d^0}{100}$ as the low values, and $100q_s$ and $100q_d$ as the high values. In Figure 4.11, the upper panels show the optimal rate, and the lower panels show the corresponding numbers of screening and treatment, respectively.

With moderate baseline scaling-up costs (middle panels), the optimal policy reallocates the intervention efforts over time in a smooth and steady way, as the quadratic cost structure could lead to significant penalties if the numbers of treatment and screening have dramatic changes over time. In particular, a relative stable number of patients, not necessarily all the diagnosed patients, receive treatment during the first 10 years. By comparing the results across different settings of q_s and q_d, we find that a higher scaling up cost results in a smoother intervention efforts throughout the horizon. When scaling-up cost is low for screening but high for treatment, the number of treatment remains steady over time, whereas the number of screening is adjusted dynamically to supply proper amount of diagnosed patients to be treated. When scaling-up cost is high for screening but low for treatment, screening tends to stay stationary, whereas the number of treatment follows the dynamic size of diagnosed population with a maximum treatment rate of 100%.

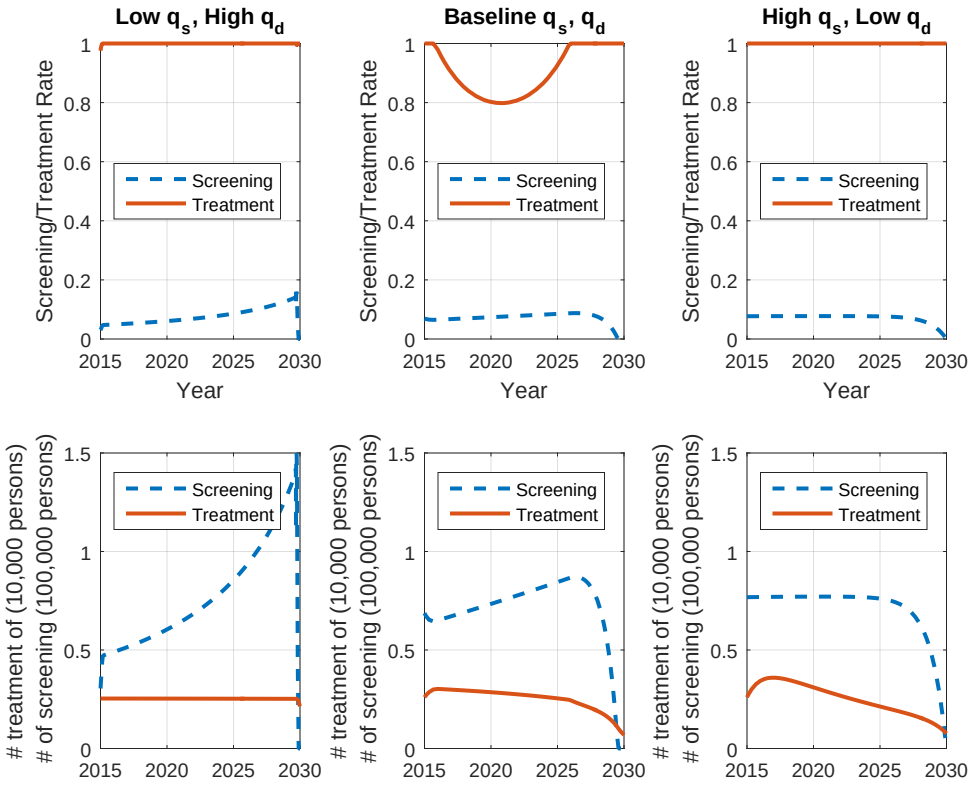


Figure 4.11: Optimal treatment and screening rates with quadratic cost functions.

Consideration of Health Outcomes for Infected People. Policy makers may also consider other possible alternatives of cost functions to seek to balance the cost with health outcomes. For example, to balance the number of liver-related deaths, we can introduce the penalty cost c_m for each liver-related death along with the intervention costs:

$$\min \int_0^{t_f} [(c_s^a(S+I+R)v + c_s^r Iv + c_d Dw) + c_m(m-\mu)(I+D)]dt;$$

or to balance the total cost with the total quality adjusted life years of the population, one can instead maximize the net monetary benefit (NMB) (208) to identify the most cost-effective policy:

$$\min \int_0^{t_f} \{\pi(e_S S + e_I I + e_D D + e_R R) - [c_s^a(S+I+R)v + c_s^r Iv + c_d Dw]\}\, dt,$$

where π represents the willingness-to-pay value, e_S, e_I, e_D, and e_R represent the quality of life for each compartment, respectively.

We remark that as long as the health outcome terms added to the objective functions are linear with respect to the control variables, the structure of the optimal control policy will not change: singular arcs may exist when both screening and treatment are control variables, and the optimal treatment policy is a pure bang-bang policy when the screening rate is fixed.

4.4.5 Constant screening rate assumption revisited

We close this section by revisiting the assumption of the constant screening rate. Our previous results have shown that the constant screening rate assumption makes the problem tractable, and also lead to the simple structure of the optimal treatment policy. Of course, such a simplifying assumption comes at the price of a suboptimal cost compared to the optimal solution without the assumption. In the following, we attempt to numerically characterize how much we may lose by assuming the constant screening rate.

Since the ideal benchmark for comparison—the exact optimal policy to the original problem (4.10)—is not available, we instead use a series of feasible solutions that allow certain changes of screening rate over time in our comparisons (Table 4.4). In particular, we obtain (1) the optimal screening and treatment policies for quadratic cost functions with various scaling-up cost values, and (2) the optimal treatment policy with a piece-wise constant screening rate where different screening rates are fixed for the first and the second half of time horizon, respectively (all combinations of a grid search for values from 0 to 15% by every 1% increment are evaluated).

The lowest total cost is attained with the optimal solution to quadratic cost function with low scaling-up cost ($q_s = \frac{q_s^0}{100},\ q_d = \frac{q_d^0}{100}$). We use this best available policy as the baseline for comparisons. Minimal constant screening rate policy results in a

Table 4.4: Comparisons of policies with dynamic screening rates.

Policy	Total cost ($ Million)	Gap
Quadratic cost function with scaling-up cost (q_s, q_d)		
(Low, Low)[1]	18.14	–
(Low, Mid)	18.34	1.1%
(Low, High)	18.82	3.8%
(Mid, Low)	18.27	0.7%
(Mid, Mid)	18.36	1.2%
(Mid, High)	18.88	4.1%
(High, Low)	18.37	1.3%
(High, Mid)	18.37	1.3%
(High, High)	18.50	2.0%
Piece-wise constant screening $(v_{[0,t_f/2]}, v_{[t_f/2,t_f]})$		
(0.13,0.01)	18.29	0.8%
(0.14,0)	18.29	0.8%
(0.13,0.02)	18.63	2.7%
(0.12,0.03)	18.64	2.7%
(0.07,0.08)	18.66	2.9%
(0.08,0.07)	18.67	2.9%
(0.09,0.06)	18.68	3.0%
(0.10,0.05)	18.69	3.0%
(0.15,0)	18.69	3.0%
(0.11,0.04)	18.69	3.0%
Constant screening rate		
$v^c = \underline{v}^c = 7.4\%$	18.58	2.4%

[1] We use the low, medium, and high values for scaling-up costs defined in Section 4.4.4.4.

2.4% gap compared with the baseline policy. Most policies with quadratic scaling-up costs outperform the constant screening rate policy, which take the advantage of a fully dynamic structure of the controls. Among the piece-wise constant screening policies as a class of more practical policies, the minimal constant screening rate policy is outperformed by only two piece-wise policies with a small gap, implying that a simple relaxation on the constant screening assumption will not significantly improve the outcomes.

A common feature among the policies with lower cost is that screening tends to be at a higher rate in early periods and reduces to a lower level or even stops later,

which indicates a simple structure of screening rate for potential improvements if the implementation is possible in practice.

To summarize, our comparisons imply that the assumption of constant screening rate does not result in a significant loss of performance. The minimal constant screening rate policy performs well and close to the best available policy with both dynamic screening and treatment rates.

4.5 Optimal Resource Allocation: Budget Constraint

In this section, we aim to address a different policy question in HCV elimination: *for a given amount of budget per year, what is the optimal allocation for screening and treatment to minimize the disease burden?* In particular, we refer to the disease burden as the cumulative number of infected people, or the cumulative prevalence. Then, the optimal control problem can be formulated as follows.

$$\begin{aligned}
\text{(P-BudgetAlloc)} \quad \min_{v,w} \quad & \int_0^{t_f} (I+D)dt && (4.25)\\
s.t. \quad & (4.1)-(4.6), \\
& c_s^a(S+I+R)v + c_s^r Iv + c_d Dw \le B && (4.26)\\
& v(t), w(t) \in [0,1],\ \forall t \in [0,t_f],
\end{aligned}$$

where B represents the annual budget which is assumed to be constant over time.

In addition to the switched roles of disease burden (prevalence) and the cost for the objective function and constraint in the formulation, the budget allocation problem (4.25) differs from the cost minimization problem (4.10) in the way how resources are allocated over time. Cost-minimizing policy employs a dynamic allocation of resources in screening and treatment, which is informed by the optimal solution but not policy makers; whereas in the budget allocation problem, the total resources

spent for screening and treatment is limited to an upper bound, which can be predetermined and planned by policy makers. As such, the budget allocation model is suitable for the cases where there is a central planner designing the HCV epidemic control program under certain budget limit.

The following proposition shows the structure of the optimal allocation strategy when a stationary size of total population is considered—a simplifying assumption for the convenience of analyses. We believe such an assumption is not restrictive as the demographic dynamics (e.g., the change of the total number of the population) are presenting limited changes compared to the epidemic dynamics (e.g., the change of the number of infected, diagnosed, or the new infections).

Proposition 4.11 (Treatment-first policy) *Consider a population with a stationary size by assuming* $\Lambda = \mu N_0$ *and* $m = \mu$*. The optimal budget allocation policy always prioritizes treatment; when there are remaining budget after covering the treatment for all diagnosed patients, allocate the remaining budget for screening.*

The treatment-first allocation policy structure is intuitive. Recall the objective is to reduce the number of the infected people, including the undiagnosed and the diagnosed, over time. Treatment directly reduces the number of infections; whereas screening takes effect in an indirect way: it redistributes the undiagnosed and the diagnosed patients within the infected population, which only matters when the budget is able to treat more patients than all the currently diagnosed patients. Without screening, excessive treatment capacity cannot reach the infected but undiagnosed population. In other words, the role of screening is intermediary by making treatment possible but not direct, and the optimal strategy is to allocate budget for screening such that number of treatment is always maximized.

Remark 4.2 *The treatment-first policy structure also holds true if the objective function in problem* (4.25) *is to minimize the final prevalence, or a convex combination*

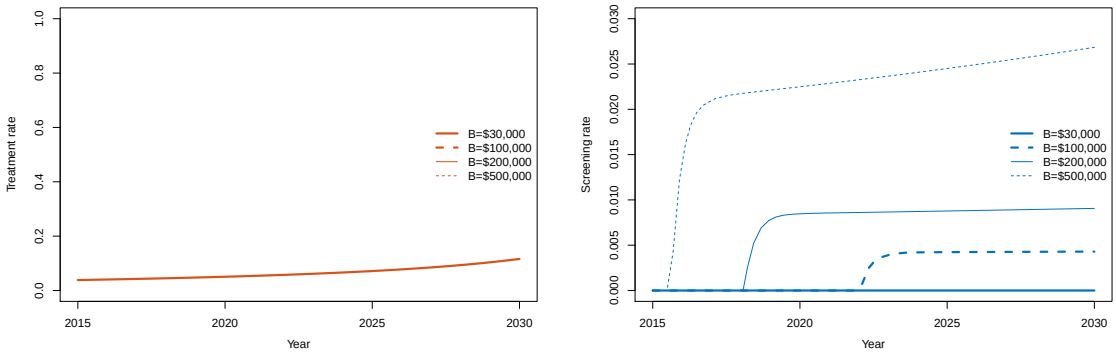
of both.

4.5.1 Numerical results

In this section, we present the numerical analyses of the budget allocation problem. We apply the same parameter values that are described earlier in Section 4.4.3 in all our numerical analyses. In the following, we will first present the optimal treatment and screening policies (solved by the GPOPS-II solver) under various annual budget, and compare the optimal allocation policy with a set of heuristic allocation policies. We also evaluate the effect of budget on the WHO targets. Finally, we explore a specific setting where the treatment-first policy might fail, and highlight the the effect of awareness of infection on the budget allocation.

4.5.1.1 Prioritize treatment

Figure 4.12 presents the optimal treatment and screening rates under various annual budget levels. The solutions clearly demonstrate a common"treatment-first" policy: allocate budget for treatment as the first priority; whenever there is remaining budget, allocate them for screening. In particular, when the annual budget is very tight (e.g., $B = \$30,000$), all efforts are devoted into treatment (Figure 4.13). When there are more budget available (e.g., $B = \$0.1 \sim 0.5M$), all budget will be first spent on treatment in early periods, but not all diagnosed patients can be treated at this stage (i.e., $w(t) < 1$); as treatment continues, there will be less diagnosed patients waiting, and then the budget is able to cover all these patients at a treatment rate $w(t) = 1$, and utilizes the remaining budget for screening. For example, in Figure 4.13, with annual budget $B = \$0.1M$, screening is available only in the second half of the planning horizon. When a higher budget is available, it is the best to treat the existing diagnosed patients as much as one can, and in the mean time to screen the population with the remaining budget.



(A) Treatment rate (B) Screening rate

Figure 4.12: Optimal budget allocation policy.

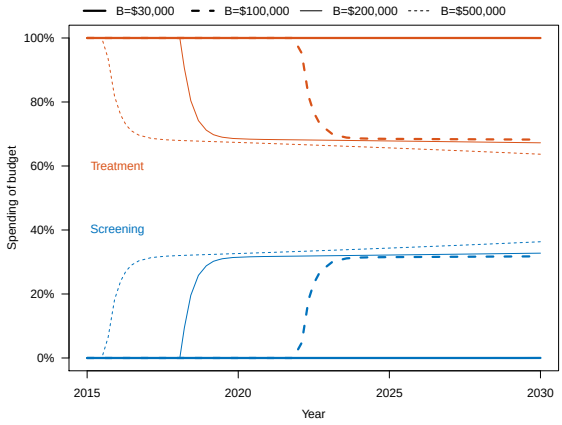


Figure 4.13: Fraction of budget allocated to screening and treatment.

4.5.1.2 Comparison of different budget allocation policies

From a practical viewpoint, a natural question for the optimal treatment-first policy is that *how much would the model outcomes depend on such a prioritization scheme?* In other words, we are interested in investigating what if the prioritization is relaxed in the resource allocation. In the following, we compare the treatment-first policy with a set of heuristic policies to seek answers to the above questions.

We define a class of policies denoted as "Scr-x policies" as follows. First allocate the *screening fraction*, $x\%$, of budget for screening, and $(100 - x)\%$ for treatment; if treatment budget is sufficiently large to cover all diagnosed patients, the remaining

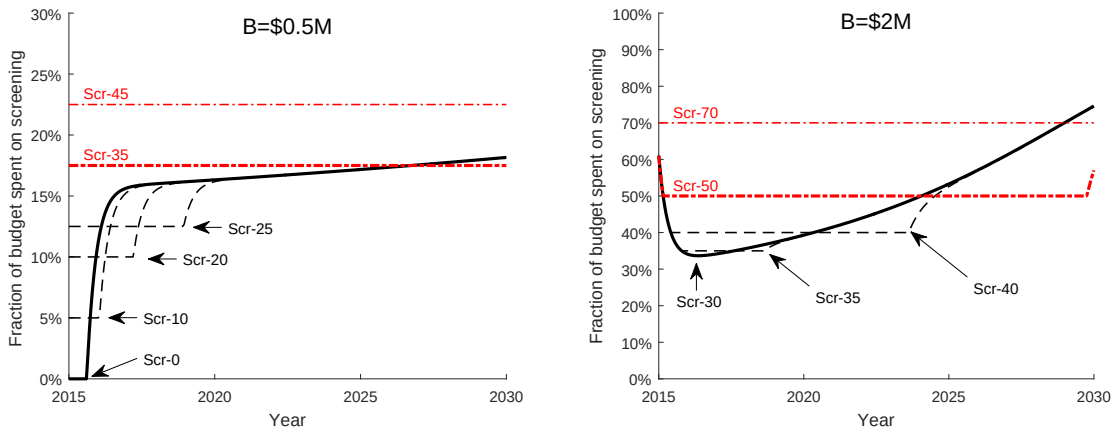
budget will be allocated for screening. That is, Scr-0 policy is synonymous with the optimal treatment-first policy. With a high budget, Scr-x policies with different value of x may represent the same policy. For example, when $B = \$2M$, Scr-0 and Scr-30 represent the same allocation policy since 70% of the budget is always sufficient to cover all the existing diagnosed patients with a treatment rate 100%.

Table 4.5: Comparison of budget allocation policies.

Allocation policy	Cumulative disease burden (million person-years)	Reduction of infected #[1]	Final diagnosed%	Final treated%
Low budget level ($0.5M/year)				
Scr-0 (treatment-first)	6.63	71%	3%	93%
Scr-10	6.63	71%	3%	93%
Scr-20	6.64	71%	3%	93%
Scr-25	6.64	71%	3%	93%
Scr-30	6.66	71%	3%	93%
Scr-35	6.74	73%	7%	83%
Scr-40	6.82	75%	14%	70%
Scr-50	6.99	79%	26%	50%
Scr-60	7.16	83%	37%	35%
Scr-70	7.33	88%	45%	24%
High budget level ($2M/year)				
Scr-30 (treatment-first)[2]	3.93	12%	28%	96%
Scr-35	3.94	12%	28%	96%
Scr-40	4.00	12%	28%	96%
Scr-45	4.18	12%	28%	96%
Scr-50	4.47	14%	39%	94%
Scr-55	4.80	23%	68%	83%
Scr-60	5.14	31%	80%	72%
Scr-65	5.47	40%	86%	62%
Scr-70	5.81	49%	90%	53%

[1] Similar results for number of new infections. [2] Identical to all Scr-x policies with $x \leq 30$.

Table 4.5 summarizes the performance of the optimal treatment-first policy and the heuristic Scr-x policies. Figure 4.14 presents the fractions of budget that are actually spent on screening in a selected subset of the policies for comparison. We find that there exists a range of screening fractions that are insensitive to the performance outcomes, such as screening fractions $\leq$ 25% for a low budget $B = \$0.5M$ and screening fractions $\leq 40\%$ for a high budget $B = \$2M$. From Figure 4.14, we observe that in these policies, since screening initially takes up certain share of budget from



(A) Annual budget of $0.5M (B) Annual budget of $2M

Figure 4.14: Fraction of budget spent on screening by different allocation policies.

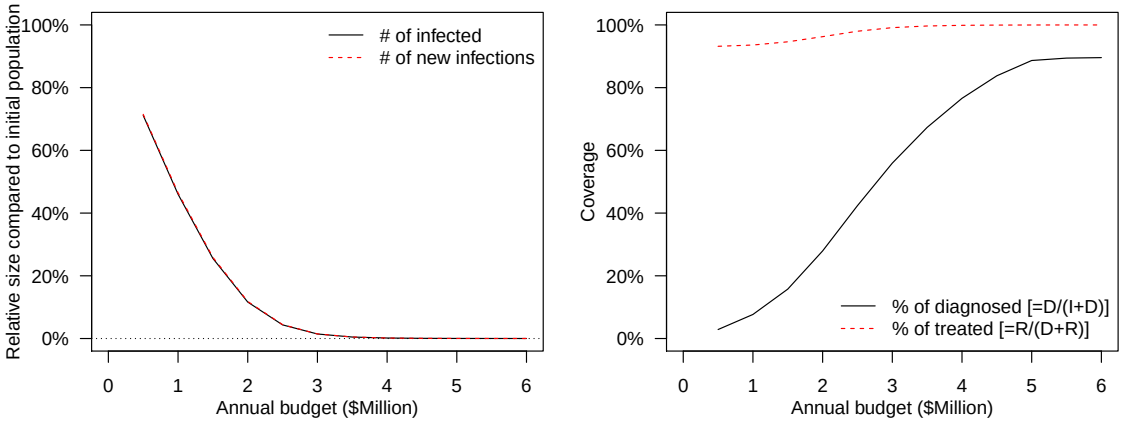
treatment, treatment needs longer time but is still able to reduce the number of diagnosed patients to a "full coverage", and then continues with maximum treatment rate 100%. Thus, diverting budget to screening within certain amount and in early periods has limited effects on the cumulative disease burden and the final outcomes, as long as the resources for treatment are able to cover all diagnosed patients in later periods. However, when too much share of budget is allocated for screening, treatment may never catch up with all diagnosed patients, leading to a poorer performance outcomes.

4.5.1.3 Budget and WHO targets

To evaluate the effect of budget on the actual epidemic control impact, we calculate the four metrics in the WHO target (215) (number of new infections, number of liver-related deaths, diagnosis rate, and treatment rate in treatment-eligible patients) for the optimal allocation policies with various budget levels.

Figure 4.15A shows the outcomes of the impact targets: the number of new infections and liver-related deaths at the final time compared with the initial values. We use the number of the infected patients $(I(t) + D(t))$ as the surrogate of number of liver-related deaths $((m-\mu)(I(t)+D(t)))$ as the two quantities have the same results

when scaled by the initial value at $t = 0$. We find that the reductions in the number of the infected and the number of new infections mostly overlap. This is because new infections $(\alpha S + \beta R)\frac{I+D}{N}$ is very sensitive to the number of the infected $(I + D)$ when the prevalence is low (e.g., <5%). Annual budget of $1.3 million reduces the number of infections by 65% of the initial value, whereas it needs a higher budget of $2.1 million annual budget to achieve the 90% target reduction of the number of new infections. We remark that a more detailed model that considers disease natural history may better differentiate the results in the number of deaths from that of new infections; yet the approximate overall estimates projected from our model are still useful to show the major impact of interventions at various budget levels.



(A) Impact targets

(B) Coverage targets

Figure 4.15: Impact of annual budget on the WHO impact and coverage targets.

Figure 4.15B shows the coverage of diagnosis (D) among the infected $(I + D)$ and the coverage of treatment (R) among the treatment-eligible patients $(D + R)$. Since budget allocation always prioritizes treatment, the treatment coverage remains high at all annual budget. On the other hand, screening is limited or even not available under very low annual budget. With annual budget of $2 million, diagnosis rate can only be as high as 28%. We remark that although diagnosis rate tends to increase with the annual budget, the diagnosis rate estimate becomes less informative given

the high budget (e.g, >$3 million), as the total number of the infected is already very low (see Figure 4.15A).

The above observations imply that if all the four WHO targets for HCV elimination are considered as constraints, these constraints may not be binding at the same time at the optimal solution. Given the low priority of screening in the optimal budget allocation policy, diagnosis rate target is difficult to achieve. Policy makers must make trade-offs between this target and the others performance measures related to disease burden (e.g., prevalence) as in the impact targets.

4.5.1.4 Budget allocation when awareness matters

As discussed earlier in Section 4.5.1.1, screening does not directly contribute to reducing disease prevalence. Its contribution is indirect, by identifying more patients eligible for treatment, especially when budget for treatment is capable of treating more diagnosed patients.

We remark that such an interpretation is subject to our current model formulation. The infection force is defined as $\alpha \cdot \frac{I+D}{N}$ (and similarly $\beta \cdot \frac{I+D}{N}$ for reinfections), implying that the numbers of the undiagnosed (in compartment I) and the diagnosed population (in compartment D) have the same effect. In other words, awareness of infection does not affect the infectivity of diagnosed patients. On the other hand, in many realistic cases of epidemic control, it is possible to reduce the chance of transmitting the disease from those who are aware of their infections through preventive interventions such as education or risk reduction programs. As such, by increasing the diagnosis rate among the infected patients, screening can reduce the infection force and thus also directly contributes to reducing the total disease burden. To capture such a phenomenon, the infection force can be modified as $\alpha \cdot \frac{I+\kappa_d D}{N}$ (and similarly $\beta \cdot \frac{I+\kappa_d D}{N}$ for reinfections), where κ_d represents the relative infectivity of the diagnosed patients.

We solve the model with the new definition of the infection force using the base case parameter settings (for India population profile), and find that κ_d varying from 0.1 to 0.9 has no effect on the optimal policy. This is possibly because the benefit of screening is still limited given that the prevalence and infection rate are not high enough such that the screening can make a difference. We next experiment several parameter settings with higher prevalence and infection rate. Figure 4.16 shows the optimal policies of allocating annual budget $B = \$2M$ for a population with an initial prevalence rate of 10%, a high infection rate $\alpha = 0.05$, and decreasing κ_d values. In such a high risk population, when awareness of infection reduces the infectivity by 70% ($\kappa_d = 0.3$), screening shows its effects: it is optimal to start with spending all budget in screening rather than treatment. The lower the κ_d value is, the greater benefits the screening will have, and thus the longer the time periods for prioritizing screening will last for. This way, the initial investment in screening reduces the infection force, which will also decrease the number of future infections as a long term benefit. Following the upfront screening periods, the optimal budget allocation resumes the treatment-first policy through the rest of time horizon.

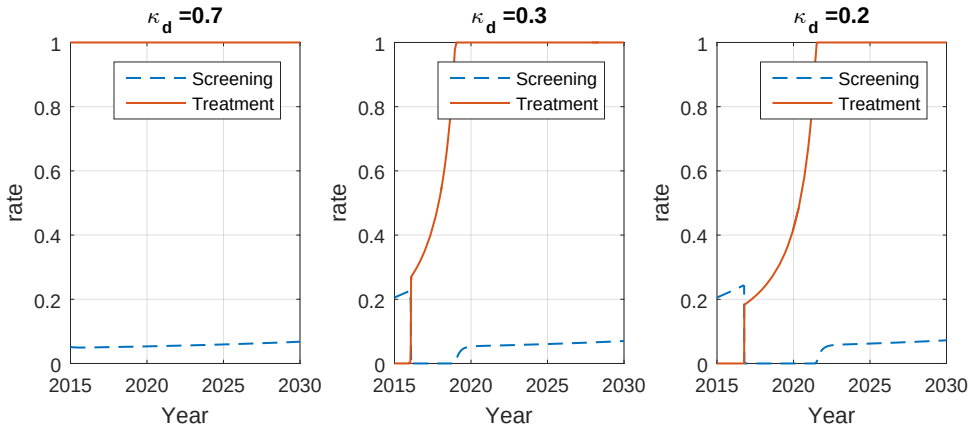


Figure 4.16: Optimal budget allocation when relative infectivity due to awareness is considered.

4.6 Discussion

Globally, 130-170 million people are chronically infected with HCV, and more than 350,000 people die from HCV every year. The recent availability of highly-efficacious therapies provide a hope to eliminate HCV by 2030 (35). However, several barriers exist before HCV can be eliminated worldwide. Our study aims to identify optimal screening and treatment policies under constrained budgetary resources, and provide insights to tackle the ongoing HCV epidemic.

In this study, we develop an epidemic model for HCV transmission. Building upon this dynamic system, we formulate two optimal control problems to identify the optimal screening and treatment policies for HCV elimination. In particular, in the first problem, we consider a cost minimization problem subject to a target prevalence constraint. We analytically show that the optimal treatment is a pure bang-bang policy without singular arcs for any given screening rate. The policy with the minimal screening rate and non-stopping treatment at the maximum treatment rate is deemed an appealing strategy, which performs closely well compared with other policies with more dynamic structures. In the meantime, it maintains a simple policy structure which is straightforward for implementation. In the second problem, we consider a budget allocation problem that aims to minimize the disease burden subject to a fixed budget constraint. We show that optimal allocation policy follows a simple treatment-first rule in both analytical and numerical results.

Our study has limitations. First, in our formulation, we do not consider behavioral and preventive interventions that could reduce HCV transmission such as harm reduction programs for high risk population (e.g., needle exchange programs, opiate substitute treatment (227)), mainly because of the lack of data to quantify the relations between the cost and effect (e.g., reduction in the infection rate). On the other hand, if the costs and effects of such interventions are available for a particular population or in a specific region, our model can be used to determine the optimal

screening and treatment strategies for the given level of such preventive intervention.

Second, our model has limited details in modeling HCV natural history. We choose to keep the model parsimonious to maintain the basic disease transmission structure, at the price of losing certain details of disease progression. As a result, our model may oversimplify the evaluation of the number of liver-related deaths as one WHO target measure, which could be dependent on specific stage of liver disease. Incorporating the information of age-groups and quality of life are also limited by this type of compartment models.

Although we present the numerical analyses as a case study for HCV epidemic elimination in India, our work can be easily extended to other countries or different HCV epidemic settings. Decision support tools can be built on the analytical policy structure and numerical solutions for policy makers to design intervention plans for HCV elimination. Recommended policies can be evaluated in a separate model so that a more comprehensive set of outcomes (e.g., deaths by liver disease stage, disability-adjusted life years for measuring disease burden) could also be provided to policy makers for a better planning.

APPENDIX A

APPENDIX FOR CHAPTER 2

A.1 Estimates of Annual CLL Incidences in the Base Case

In the simulation model, the annual incidence of CLL between 2000–2011 were obtained through model calibration (see Appendix A.2), and the annual incidence between 2011-2014 were obtained from published sources (196, 198, 199, 197). From 2015 onwards, we assumed stationary annual incidences in our base case results, i.e., a constant annual incidence as the same as the 2014 estimate (Table A.1). In the sensitivity analysis, we adjusted for increase in the annual CLL incidence because of the aging of US population (see Appendices A.6 and A.7.3).

Table A.1: Annual incidence estimates.

Year	Annual incidence
2000-2010	SEER incidence rescaled in calibration
2011	14,570
2012	16,060
2013	15,680
2014	15,720
2015	15,720
2016	15,720
2017	15,720
2018	15,720
2019	15,720
2020	15,720
2021	15,720
2022	15,720
2023	15,720
2024	15,720
2025	15,720
2026	15,720
2027	15,720
2028	15,720
2029	15,720
2030	15,720

A.2 Model Calibration

We developed a microsimulation model, simCLL (simulation model of CLL management), that simulated the clinical practice of chronic lymphocytic leukemia (CLL) management in the US from 2011 to 2025. Model components included population characteristics, incidence of CLL, survival outcomes using existing and emerging therapies, cost of treatment and disease management, and quality of life associated with CLL.

We started the simCLL from year 2000 and matched simulation outcomes with Surveillance, Epidemiology, and End Results (SEER) data until 2010. The period from 2000-2010 was used as warm-up phase in our simulation model. The purpose of the warm-up phase was to generate a reasonable population profile that closely matched the observation data from SEER so that the base case simulation will start with this baseline population from 2011. In the warm-up phase, if the simulation outcomes did not match with SEER data, we calibrated (i.e., adjusted) model input parameters until the model outcomes and SEER data matched closely.

In particular, during the 2000-2010 warm-up period, the simulation model started from an empty population in 2000, generated new incidences for each 5-year age group (i.e., 30-34, 35-39, etc.) in each model cycle according to the SEER incidence data (211). The following input parameters were calibrated: time-to-treatment probability, progression and mortality risks for fit and unfit patients, mortality risks for relapsed patients. The primary calibration target was 11-year CLL prevalence and the age distribution in 2011 (i.e., at the end of warm-up period). We identified 100 best combinations of calibration parameters by grid search, and used their averages as the calibrated values. Figure A.1 shows the comparison of CLL prevalence in 2011 between the simulated population and the SEER data (Figure A.1-A), and age distributions of CLL patients in 2011 (Figure A.1-B). Therefore, with the calibrated parameters, we obtained the population characteristics (i.e., joint distribution of age

and health states) in 2011, which was used as the initial population for our base case results (Figure A.2).

Note that the SEER data does not fully cover the US population. To recover the total disease population in the US, we scaled up the number of new CLL incidences during 2000-2011 in our simulation model such that the total number of CLL patients in the simulation is close to the national estimates of 126,299 cases in 2011 (210).

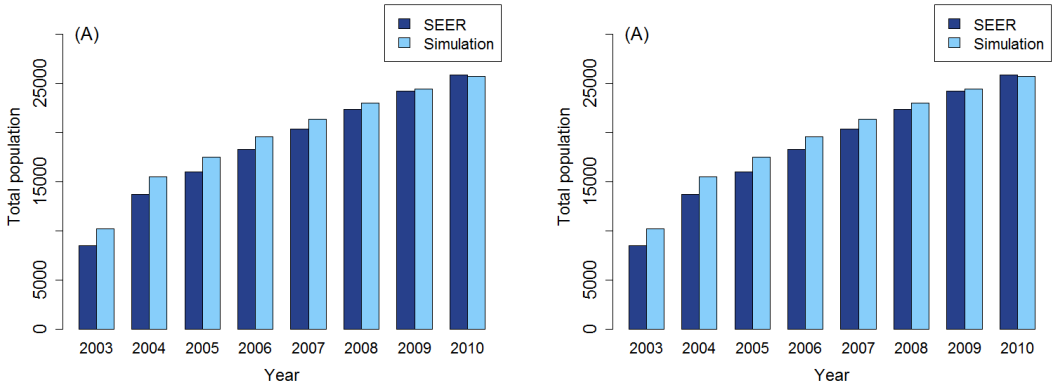


Figure A.1: Model calibration results.

(A) Comparison of limited-duration prevalence from SEER and simulation by year.
(B) Comparison of age distribution at the end of calibration phase.

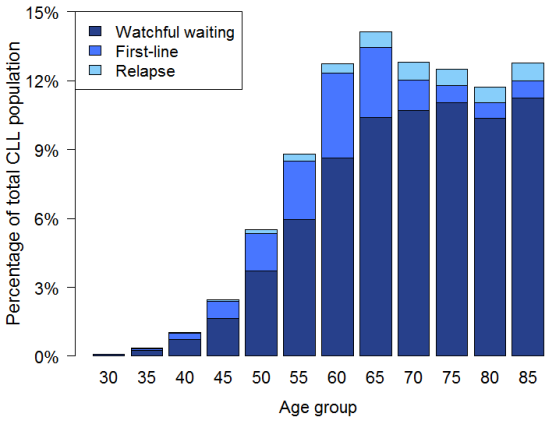


Figure A.2: Calibrated population distribution by age and health states in 2011.

A.3 Survival Model Fittings, State Transition Probability Estimation and Validation

Our microsimulation model, simCLL, included four health states: watchful waiting, first-line treatment, relapse, and dead (Figure A.3). The transition probabilities between the health states were estimated from survival distribution from the published survival curves. In particular,

- Transition probabilities from watchful waiting to first-line treatment were based on time-to-treatment results from a large clinical study (171);
- Transition probabilities from first-line treatment to relapse and death states were based on the progression-free survival (PFS) and overall survival (OS) distributions for the specific first-line treatment option and patient characteristics;
- Transition probabilities from relapse to death were based on the OS distribution of treatment for relapsed patients.

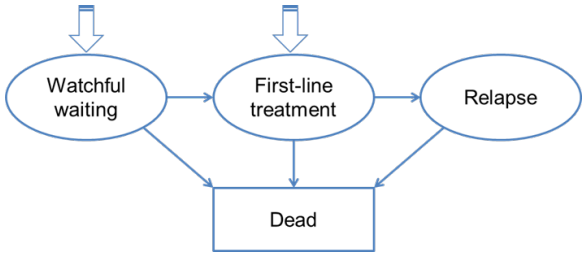


Figure A.3: Clinical pathways of CLL patients in microsimulation model.

A.3.1 Fitting of survival distributions

We first digitized the Kaplan-Meier curves from published studies (1), and fitted the extracted data points with parametric survival models. We considered the following commonly used parametric survival models in survival analysis:

- Exponential distribution: $S(t) = \exp(-\lambda t)$ (shape parameter λ).

- Weibull distribution: $S(t) = \exp((-\lambda t)^p)$ (shape parameter λ, scale parameter p).
- Log-logistic distribution: $S(t) = \frac{1}{1+(\lambda t)^{1/\gamma}}$ (shape parameter λ, scale parameter γ).
- Log-normal distribution: $S(t) = 1 - \Phi(\frac{\ln t - \mu}{\sigma})$ (shape parameter μ, scale parameter σ, the cumulative distribution function Φ of the standard normal distribution).
- Gompertz distribution: $S(t) = \exp\left(-\lambda\gamma^{-1}(\exp(\gamma t) - 1)\right)$ (shape parameter λ, scale parameter γ).

The best fitted parametric survival distributions were selected based on the goodness-of-fit metrics (e.g., Akaike information criterion, Bayesian information criteria), as well as visual inspections for a reasonable extrapolation beyond the limited follow-up periods (151). Parameters of the fitted distribution for each survival model are provided in Table A.2, and comparisons of goodness-of-fit between survival distributions are provided in Table A.3.

Table A.2: Fitted survival distributions and parameters.

Treatment	Survival	Survival model	Shape parameter	Scale parameter
Time to treatment				
Age $\geq$ 55 yr		weibull	0.011	0.742
Age $<$ 55 yr		weibull	0.014	0.708
First-line				
FCR, fit	OS	weibull	0.006	1.202
FCR, fit	PFS	weibull	0.015	1.349
FCR, del17p	OS	log-normal	3.452	0.824
FCR, del17p	PFS	log-normal	2.53	0.629
Clb, unfit	OS	weibull	0.01	1.164
Clb, unfit	PFS	gompertz	0.047	-0.013
Clb, del17p	OS	weibull	Extrapolated[a]	
Clb, del17p	PFS	gompertz	Extrapolated[a]	
GClb, unfit	OS	weibull	0.012	1.751
GClb, unfit	PFS	weibull	0.031	1.959
GClb, del17p	OS	weibull	Extrapolated[a]	
GClb, del17p	PFS	weibull	Extrapolated[a]	
Ibrutinib, fit/unfit	OS	log-normal	Extrapolated[b]	
Ibrutinib, fit/unfit	PFS	log-normal	4.377	1.229
Ibrutinib, del(17p)	OS	exponential	0.007	
Ibrutinib, del(17p)	PFS	exponential	0.004	
Relapse				
FCR	OS	weibull	0.015	0.852
BR	OS	weibull	0.023	1.451
Idelalisib+R	OS	weibull	0.004	0.867
Ofatumumab	OS	log-logistic	0.071	0.745
Ibrutinib	OS	log-normal	5.039	2.006
HSCT	OS	log-normal	4.069	2.224

a: For patients with del(17p), PFS and OS outcomes for Clb treatment were extrapolated by applying the hazard ratio (HR) associated with del(17p) (i.e., HR=9.3 for OS, and 7.3 for PFS) observed in CLL8 trial (95).
b: Due to limited events observed in the ibrutinib arm of RESONATE2 trial, we did not directly fit the survival model to the Kaplan-Meier estimates of ibrutinib arm; we first fitted the survival model to the Clb arm (with a log-normal distribution with a shape parameter 5.119 and a scale parameter 1.938) and then applied the estimated hazard ratio associated the ibrutinib arm (HR=0.16) (36) to the fitted distribution.

Table A.3: Comparison of goodness-of-fit for different parametric survival distributions.

		AIC					BIC				
		Expo -nential	Weibull	Log -logistic	Gom -pertz	Log -normal	Expo -nential	Weibull	Log -logistic	Gom -pertz	Log -normal
Time to treatment											
Age 55yr		-290.6	-451.1	-354.0	-361.2	-343.7	-285.6	-443.5	-346.4	-353.6	-336.1
Age <55yr		-252.6	-392.8	-306.1	-307.4	-299.1	-247.6	-385.2	-298.6	-299.9	-291.6
Firstline											
FCR, fit	OS	-272.7	-291.1	-285.2	-305.8	-271.4	-269.0	-285.5	-279.6	-300.2	-265.9
FCR, fit	PFS	-238.2	-370.2	-330.4	-375.8	-307.3	-233.9	-363.8	-324.0	-369.4	-300.8
FCR, del17p	OS	-106.0	-137.7	-153.3	-123.3	-158.7	-102.2	-132.0	-147.5	-117.6	-153.0
FCR, del17p	PFS	-50.1	-81.7	-99.6	-67.2	-100.1	-46.8	-76.9	-94.8	-62.4	-95.2
Clb, unfit	OS	-341.9	-351.7	-328.7	-375.9	-312.7	-336.8	-344.0	-321.0	-368.2	-305.1
Clb, unfit	PFS	-417.8	-497.1	-530.8	-545.4	-530.9	-412.3	-488.9	-522.6	-537.2	-522.7
GClb, unfit	OS	-351.4	-380.4	-376.7	-404.1	-366.3	-346.5	-373.0	-369.4	-396.8	-359.0
GClb, unfit	PFS	-132.5	-356.7	-328.7	-307.6	-308.9	-128.2	-350.3	-322.3	-301.1	-302.5
Clb, unfit*	OS	-526.0	-525.2	-532.0	-529.9	-548.7	-521.6	-518.6	-525.4	-523.2	-542.0
Ibrutinib, unfit	PFS	-340.9	-427.2	-428.6	-413.2	-427.7	-336.6	-420.8	-422.2	-406.8	-421.3
Relapse											
FCR	OS	-403.2	-486.1	-481.0	-479.4	-470.5	-398.4	-478.8	-473.7	-472.1	-463.2
BR	OS	-282.0	-332.6	-307.5	-370.6	-293.3	-276.8	-324.7	-299.7	-362.7	-285.5
Idelalisib+R	OS	-668.3	-671.7	-672.9	-681.0	-679.8	-662.8	-663.4	-664.6	-672.8	-671.5
Ofatumumab	OS	-508.8	-521.6	-546.3	-512.5	-541.9	-503.4	-513.5	-538.2	-504.4	-533.8
Ibrutinib	OS	-462.0	-504.9	-510.4	-516.8	-512.8	-457.1	-497.6	-503.1	-509.5	-505.5
HSCT	OS	-179.9	-322.4	-352.0	-449.8	-364.0	-175.0	-315.1	-344.7	-442.5	-356.7

AIC, Akaike information criterion; BIC, Bayesian information criteria. * Survival outcome for Clb arm in RESONATE2 trial (36)

A.3.2 Conversion to transition probabilities

We converted the fitted survival distribution to the transition probabilities for each model cycle (i.e., 4 weeks in our microsimulation) using the Bayes' Formula: given the survival function $S(t)$ for the random survival time T, the transition probability $m(t)$ for model cycle t is calculated by

$$m(t) = \frac{\mathbb{P}(t \leq T < t+1)}{\mathbb{P}(t \leq T)} = \frac{S(t) - S(t+1)}{S(t)}.$$

A.3.3 Mortality and progression risk

Mortality risk determined transition from any state to the "death" state, which was defined as the maximum value of cause-specific mortality and other-cause mortality in each cycle. The cause-specific mortality for each cycle was estimated based on OS data of each specific treatment (Table 3.3) following the above described procedure; other-cause mortality is age-dependent and was estimated from the US life tables (16). For watchful waiting state, we assigned only the background mortality.

Progression risks determined transitions from the first-line treatment state to relapse state. By definition, PFS accounted for both progression and death events during the follow-up time. Thus, to obtain the risk for progression only, we subtracted the estimated mortality risks from the transition probabilities estimated from the PFS curves.

A.3.4 Validation of model-projected survival

To validate that the simulation model predicted the survival results in line with the observed survival data of each specific treatment, we collected the progression-free time for first-line patient and the survival time (after progression) for relapse patient. For those who were alive at the end of the simulation, their progression-free time or survival time were considered as censored. Then we obtained the Kaplan-Meier estimates of the simulation survival results. In each of the following figure, we plotted

the observed survival curves obtained from published trials (in red), fitted survival curves as the model input (in blue), and the simulated survival result (in black).

Figure A.4 shows reasonable matches between the model outputs and observed survival from clinical studies. In a few relapse treatment settings, the simulated survival curves were lower than the observed survival in very late periods; this is mainly because these patients are typically in old ages where the age-specific background mortality could dominate the mortality risk extrapolated from the fitted survival curve. Overall, these results represent a good validation of our simulation model.

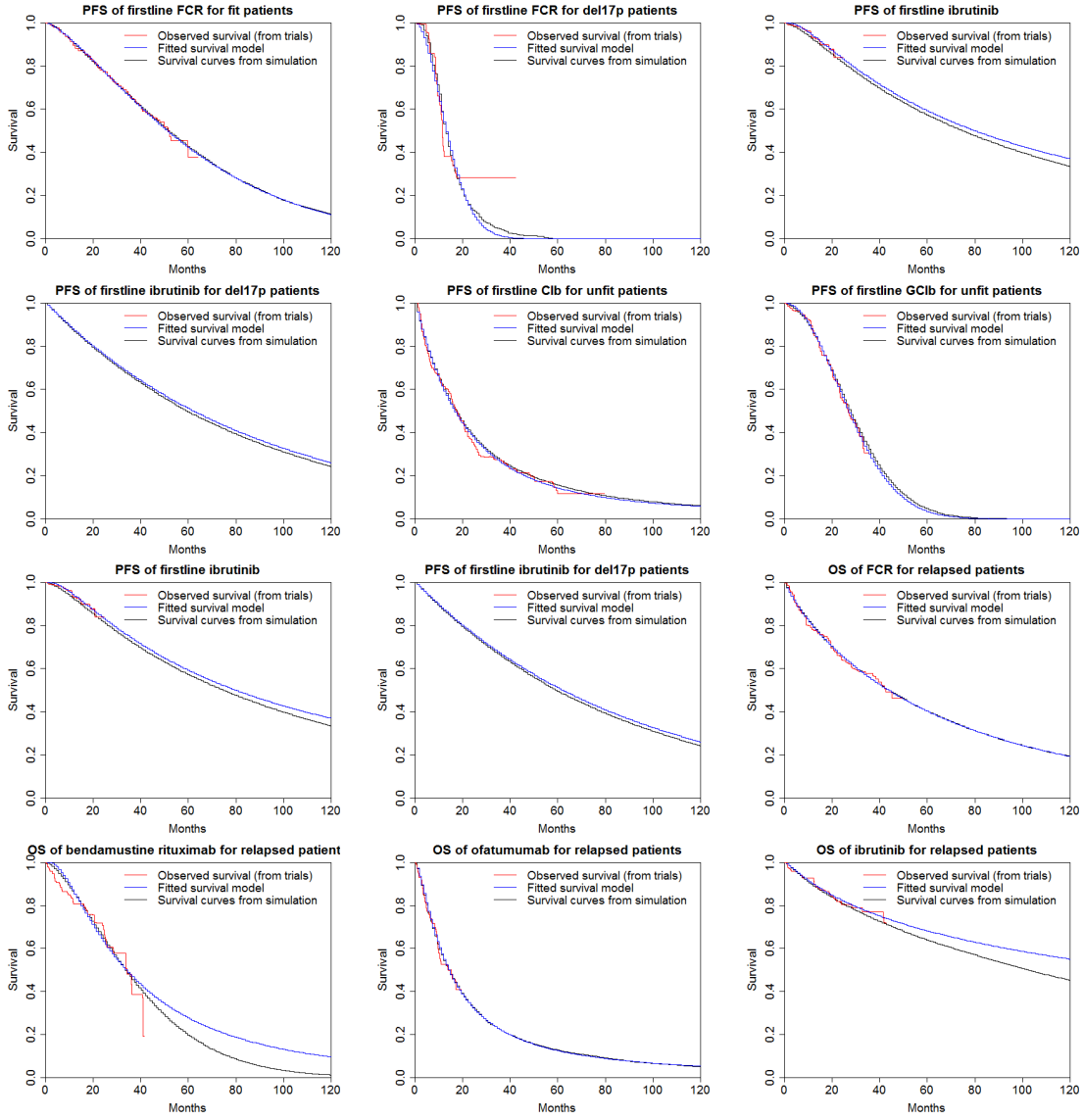


Figure A.4: Validation of survival results from the simulation model.

A.4 Cost Parameter Estimation

Table A.4: Estimation of treatment administration cost.

Treatment	Schedule of treatment administration			Hours	Cost per administration [a] ($)	Cost per cycle [b] ($)
FCR	cycle 1	day 1	R	4	220.62	817.02
		day 2,3,4	FC	2	198.8	
	cycle 2-6	day 1	FCR	5	318.23	715.83
		day 2,3	FC	2	198.8	
GClb	cycle 1	day 1, 2, 8, 15	obinutuzumab	4	220.62	882.48
	cycle 2-6	day 1	obinutuzumab	4	220.62	220.62
BR	cycle 1-6	day 1	RB	5	283.55	419.42
		day 2	B	1	135.87	
Idelalisib+rituximab		wk 1, q2wk*4, q4wk*3	R	4	220.62	661.86 (cycle 1) 275.78 (cycle 2-5) [c]
Ofatumumab		wk 1, 2	ofatumumab	7	305.37	1108 (cycle 1)
		q.wk*6, q4wk*4	ofatumumab	5	248.87	398 (cycle 2-6) [c]

a: The cost of administration for each cycle was calculated based on the total infusion hours, where the initial hour, additional hours, and switching to a different drugs have different cost rates (corresponding to CPT 96413, 96415, and 96417, respectively). For example, on day 1 of cycle 2 for FCR treatment, the total 5 hours consist of 1 initial infusion hour (CPT 96413) and 2 additional infusion hours (CPT 96415) for rituximab, and 2 additional infusions of different drugs (CPT 96417).
b: One (model) cycle represents 4 weeks.
c: For some regimens which do not repeat the same schedule every 4 weeks, we calculate the average cost per cycle.

Table A.5: Estimation of drug cost.

Treatment	Dose	Unit cost	Cost [a]
FCR			\$7455 (cycle 1), \$10063 (cycle 2-6)
Fludarabine phosphate (first 3 days)	25 mg/m2 per day	\$113.4 per 50 mg	
Cyclophosphamide (first 3 days)	250 mg/m2 per day	\$343.44 per 500 mg	
Rituximab (day 0, cycle 1)	375 mg/m2 (cycle 1)	\$869.22 per 100 mg	
Rituximab (day1, cycle 2-6)	500 mg/m2 (cycle 2-6)	\$869.22 per 100 mg	
GClb			\$19063 (cycle 1), \$6679 (cycle 2-6)
Chlorambucil (days 1 and 15)	0.5 mg/kg	\$640.32 per 100 mg	
Obinutuzumab (days 1, 8, and 15 of cycle1)	1000 mg	\$6192 per 1000 mg	
Obinutuzumab (day 1 of cycle 2-6)	1000 mg	\$6192 per 1000 mg	
Cholrambucil (days 1 and 15)	0.8 mg/kg	\$640.32 per 100 mg	\$779 (cycle 1-12)
Ibrutinib (daily)	420 mg per day	\$11002 for 30pcs	\$10270 (every 4 wks)
Bendamustine and rituximab			\$13619 (cycle 1), \$16227 (cycle 2-6)
Bendamusine (days 1 and 2)	70 mg/m2	\$1255.74 per 45 mg	
Rituximab (day 0 of cycle 1)	375 mg/m2	\$869.22 per 100 mg	
Rituximab (day 1 of cycle 2-6)	500 mg/m2	\$869.22 per 100 mg	
Idelalisib + Rituximab			
Idelalisib (twice a day)	150 mg	\$9495.36 per 60×150mg	\$8862.34 (every 4 wks)
Rituximab (5 cycles)	375 mg/m2 (first dose) 500 mg/m2 (q2wk for 4 doses and q4wk for 3 doses)	\$869.22 per 100 mg	\$23469 (cycle 1), \$10865 (cycle 2-5) [b]
Ofatumumab	300 mg for initial dose 2000 mg weekly for 7 doses, q4wk for 16 weeks	\$571.56 per 100 mg	\$36008 (cycle 1), \$18290 (cycle 2-6) [b]

a: Calculation of treatment dosing was based on an average body surface area of 1.86 m^2 and an average body weight of 76 kg (121). The cost of each drug was rounded up to the next vial. One (model) cycle represents 4 weeks.
b: For some regimens which do not repeat the same schedule every 4 weeks, we calculate the average cost per cycle.

A.5 Out-of-pocket Cost Calculation for Medicare Patients

Unlike chemotherapies which are covered by Medicare Part B, oral targeted therapies are covered in Medicare Part D. Patients in Medicare Part D would need to pay out-of-pocket expenses for these drugs. In particular, after spending the initial coverage limit, patients enter the coverage gap (known as the "donut hole"). Although the Affordable Care Act (ACA) has started to gradually lower the cost share for patients in the coverage gap to reduce the total out-of-pocket costs (expected to reduce to 25% cost sharing by 2020, see Table A.6), CLL patients may still face high out-of-pocket costs in the oral targeted agents. We considered the reducing coverage gapes when we estimated the average lifetime out-of-pocket drug costs for CLL patients.

Table A.6: Comparisons of Medicare part D standard benefit model plan parameters since 2011.

Year	Deductible	Initial coverage limit	Total covered part D drug out-of-pocket spending including coverage gap [a]	Cost sharing in the coverage gap (brand)	Cost sharing in the coverage gap (generic)
2011	310	2840	6447.5	0.5	0.93
2012	320	2930	6657.5	0.5	0.86
2013	325	2970	6733.5	0.475	0.79
2014	310	2850	6455	0.475	0.72
2015	320	2960	6680	0.45	0.65
2016	360	3310	7062.5	0.45	0.58
2017 [b]	360	3310	7062.5	0.4	0.51
2018	360	3310	7062.5	0.35	0.44
2019	360	3310	7062.5	0.3	0.37
2020 onwards	360	3310	7062.5	0.25	0.25

Note: Adapted from http://www.q1medicare.com/PartD-The-MedicarePartDOutlookAllYears.php
a: Catastrophic coverage starts after this point.
b: Assume the deductible and initial coverage limit in 2017 and onwards to be the same as those in 2016.

The out-of-pocket cost was calculated as follows (213). Patients first pay all drug spending within the deductible (e.g., $320 in 2015 plan), and have 25% cost sharing until the total drug spending has reached the initial coverage limit (e.g., $2960 in 2015 plan). Then patients enter the coverage gap, where they are responsible for a large share of drug cost (the exact share depends on the plan of a specific year, see Table A.6), until the patients' cumulative total out-of-pocket costs of this year have reached

the total out-of-pocket spending limit (e.g., $6680 in 2015 plan). The rest of drug cost will be covered by the catastrophic coverage with a 5% cost share. We provided two working examples for calculating the annual out-of-pocket costs for ibrutinib for 2015 and 2020 plan, respectively (Table A.7).

Given the annual total cost of $130,000 for ibrutinib, the annual out-of-pocket costs of ibrutinib was estimated as $10,400-$10,600, which were not affected substantially by the reducing coverage gap with ACA, because majority of the ibrutinib cost for one year was covered by the catastrophic coverage category.

Table A.7: Examples of out-of-pocket cost calculation for ibrutinib based on 2015 and 2020 Medicare Part D plan.

Coverage	Drug cost[a]	Out-of-pocket cost	Cumulative drug cost	Cumulative out-of-pocket cost
2015 Plan				
Deductible	320	320	320	320
Initial coverage	2960-320=2640	2640*0.25=660	2960	320+660=980
Coverage gap	3720/0.45=8267	6680-2960=3720	2960+8267=11227	3720+980=4700
Catastrophic coverage	130000-(320+2640+8267) =118773	5938	130000	4700+5938=10638
2020 Plan				
Deductible	360	360	360	360
Initial coverage	3310-360=2950	2950*0.25=737.5	3310	1097.5
Coverage gap	3752.5/0.25=15010	7062.5-3310 =3752.5	3310+15010=18320	3752.5+1097.5=4850
Catastrophic coverage	130000 -(360+2950+15010) =111680	5584	130000	4850+5584 =10434

a: Assuming the total cost of ibrutinib for one year is $130,000 in this example.

A.6 Adjustment of Annual CLL Incidence with Aging US Population

In the base case, we assumed a stationary annual CLL incidences from 2015 onwards (Appendix A.1). In the sensitivity analysis, we considered an increase in CLL incidence because of the aging of US population (Table A.8). Thus, we adjusted the annual incidence of CLL with the US population projection data as follows.

1. We obtained age distribution of CLL incidences in 2012 from the most recently released SEER data;

2. Given the estimate of total CLL incidences in the US in 2012 (198), we calculated the number of CLL incidences by age groups in the US in 2012.

3. Based on the 2012 US population data from US Census Bureau (223), we estimated the age group-specific incidence rates (e.g., # per 100,000 persons in age group 60-64 year-old)

4. We applied the estimated age group-specific incidence rates to the US population projection data (224) (the population projection data is available for every 5 years from 2010, so we interpolate the estimates within each 5 years).

Table A.8: Increase of population in old age groups compared with 2015 population.

Age	2020	2025	2030	2035	2040	2045	2050
60-64	11.0%	12.3%	4.5%	1.7%	4.3%	12.2%	17.2%
65-69	12.5%	25.2%	27.1%	18.6%	15.8%	19.1%	28.4%
70-74	29.1%	45.8%	63.0%	66.1%	55.8%	52.6%	57.6%
75-79	21.7%	58.2%	79.9%	102.4%	107.4%	95.8%	92.8%
80-84	9.5%	34.9%	77.2%	103.5%	131.0%	138.9%	127.7%
85-89	4.7%	14.6%	37.9%	79.9%	122.1%	164.6%	195.0%
65	16.6%	35.4%	52.1%	62.7%	69.1%	74.2%	80.8%
75	12.8%	37.7%	65.8%	95.6%	118.8%	130.0%	135.2%

Based on population projection data from US Census Bureau (224).

Following the above-described approach, we obtained the incidence estimates accounting for the aging of the US population (Table A.9). We also calculated the

percentages of new CLL incidences with age >70 and >65 by year (Table A.9), which clearly shows the aging trend in the CLL incidences in the next few decades.

Table A.9: Annual incidence estimates adjusted for aging US population.

Year	Annual incidence	% age >70	% age >65
2000-2010	SEER incidence rescaled in calibration	-	-
2011	14,570	51.1%	65.4%
2012	16,060	51.5%	67.0%
2013	16,228	51.8%	67.5%
2014	16,569	51.7%	67.7%
2015	16,912	51.5%	68.0%
2016	17,312	52.1%	68.5%
2017	17,707	52.6%	69.1%
2018	18,105	53.0%	69.6%
2019	18,505	53.5%	70.0%
2020	18,902	53.9%	70.5%
2021	19,340	54.7%	71.3%
2022	19,777	55.4%	72.0%
2023	20,216	56.1%	72.7%
2024	20,653	56.8%	73.4%
2025	21,094	57.5%	74.0%
2026	21,516	58.4%	74.6%
2027	21,938	59.3%	75.2%
2028	22,364	60.1%	75.8%
2029	22,787	60.9%	76.4%
2030	23,210	61.7%	77.0%

A.7 Results of Sensitivity Analyses

A.7.1 Analysis of drug cost discount

We used 26% discount off average wholesale price (AWP) as our base case estimates of drug cost. In sensitivity analyses, we considered different AWP discount estimates. We applied the 37% discount of AWP as lower limits of drug costs, suggested by another study by congressional budget office which estimated that the "lowest price paid by any private-sector purchaser for the drug product" (including discounts, rebates, chargebacks, and other pricing adjustments) to be 37% lower than AWP (104, 139). In addition, we performed the analyses with full AWPs as upper limits of drug costs for comparisons. Table A.10 presents the major change of the model results compared with our base case results.

The ICER of oral-targeted therapies remained greater than $160,000/QALY given the range of AWP discounts, which is still above a reasonable range of willingness-to-pay ($50,000/QALY–$150,000/QALY). These results show that even using the expected discounted drug price, these therapies will not be deemed cost-effective.

Table A.10: Comparisons of results with drug price discount.

	Base case (26% discount)	37% discount on drug cost	Change	Full AWP	Change
Annual cost of CLL management					
CIT					
2011	743,062,907	649,620,128	-13%	914,519,232	23%
2014	948,972,374	827,749,057	-13%	1,192,129,412	26%
2016	994,163,585	867,002,274	-13%	1,248,639,494	26%
2020	1,137,790,265	991,094,998	-13%	1,425,418,922	25%
2025	1,116,012,859	972,161,578	-13%	1,398,157,351	25%
OTT					
2011	743,062,907	649,620,128	-13%	914,519,232	23%
2014	917,874,069	802,374,898	-13%	1,148,186,659	25%
2016	1,657,380,404	1,427,638,370	-14%	2,159,133,619	30%
2020	4,238,284,439	3,621,650,816	-15%	5,680,660,839	34%
2025	5,130,373,316	4,381,990,566	-15%	6,892,176,447	34%
Cumulative cost 2011-2025					
CIT	15,218,203,943	13,266,836,906	-13%	19,036,395,847	25%
OTT	44,395,963,354	38,031,694,677	-14%	59,021,379,222	33%
Increment	29,177,759,410	24,764,857,771	-15%	39,984,983,375	37%
Lifetime cost of OTT					
2011	146,744	127,014	-13%	188,021	28%
2014	331,376	284,982	-14%	435,418	31%
2016	604,092	516,065	-15%	811,620	34%
Lifetime out-of-pocket cost of OTT					
2011	9,246	8,412	-9%	11,236	22%
2014	27,416	25,315	-8%	32,339	18%
2016	56,560	52,150	-8%	66,944	18%
Incremental cost-effectiveness					
Incremental total discounted cost	20,167,080,889	17,121,161,329	-15%	27,366,527,123	36%
ICER ($/QALY)	189,226	160,646	-15%	256,778	36%

A.7.2 One-way sensitivity analysis

In the one-way sensitivity analysis, we ran the model with the upper and lower limits of each parameter value. For some parameters that belong to the same category, we

changed the values of these parameters in the same group. For example, we changed the cost of all oral targeted agents by ±20% together (rather than changing the cost of ibrutinib and idelalisib separately); we adjusted the survival distributions for all oral targeted therapies with hazard ratios of 0.8 or 1.2 together (rather than adjusting the survival distributions for one therapy at each time). Utility and oral targeted therapy discountinuation probability were varied within their reported confidence interval,

We presented the sensitivity analysis results for four main model outcomes using tornado diagrams (Figure A.5): (1) *incremental* annual cost of CLL management of oral targeted therapies compared with chemoimmunotherapies in 2016, when oral targeted therapies start to be used in first-line setting; (2) *incremental* cumulative cost of oral targeted therapies from 2011–2025; (3) lifetime treatment cost of oral targeted therapies for patients using oral targeted therapies as first-line treatment after 2016. The tornado diagram for the incremental cost-effectiveness ratio of oral targeted therapies for 2011-2025 was presented in Figure 2.5A in the main text. Following tornado diagrams depict sensitive model parameters and their effects to the model outputs.

A.7.3 Scenario analysis

We evaluated the model results under two scenarios considering (1) the CLL incidence rates rising with aging population in the US (see Appendix A.6 for more details), and (2) the gradual transition from CIT to oral targeted therapies for fit patients in the first-line setting, where we assumed 25%, 50%, 75%, and 100% utilization of oral targeted therapies in 2016, 2017, 2018, and 2019, respectively. We presented the results and comparisons to the base case in Table A.11.

A.7.4 Probabilistic sensitivity analysis (PSA)

To define the sampling distribution for each parameter in the PSA, we assumed a mean value of its baseline value, and a standard deviation of half width of its range

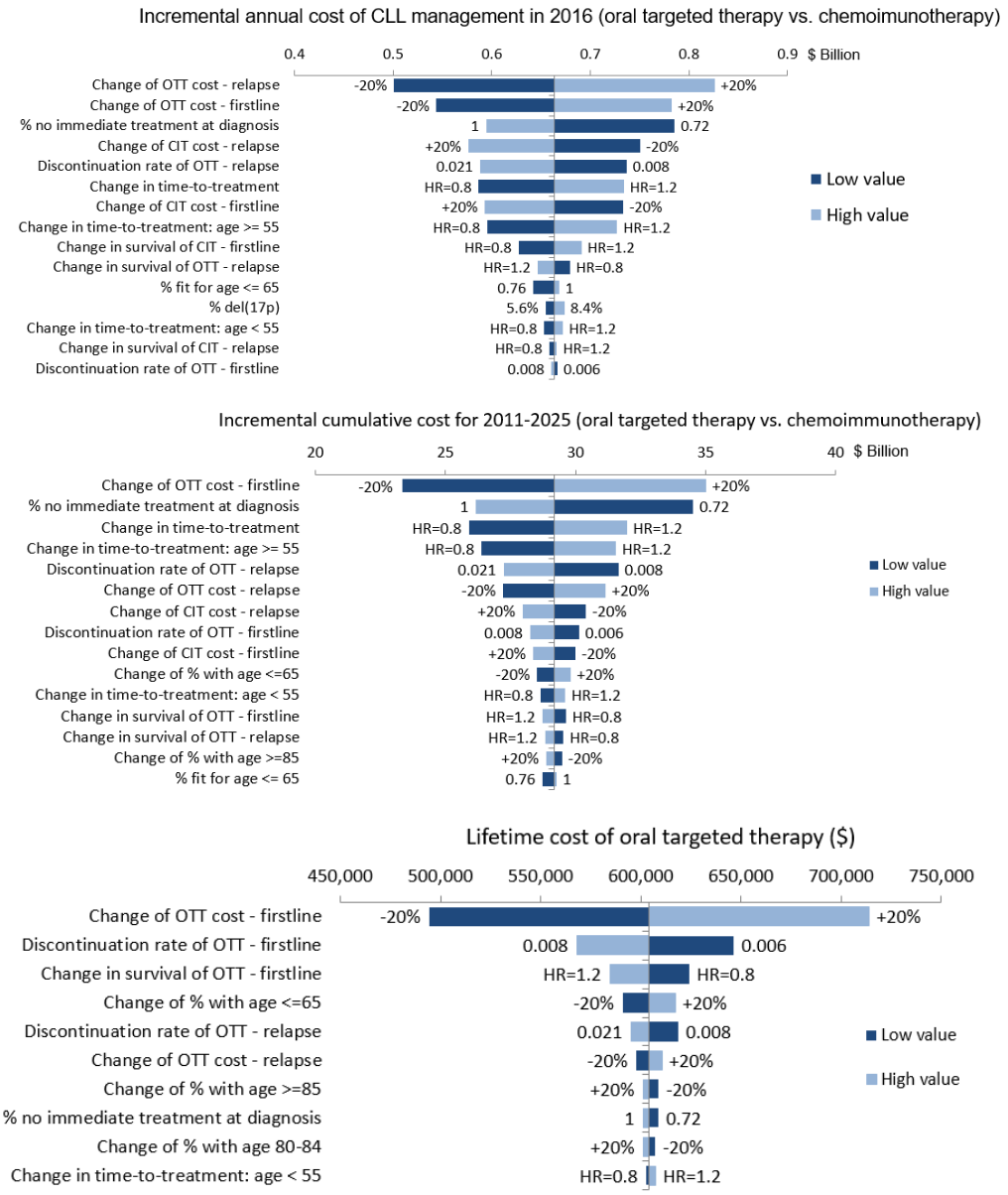


Figure A.5: Tornado diagrams for one-way sensitivity analysis.

that is defined for one-way sensitivity analysis. We follow the recommended statistical distributions by the parameter types (31), which were defined by: gamma distribution for cost parameters, binomial distribution for probability and utility parameters, lognormal distribution for hazard ratio parameters. We run the PSA for 1000 iterations.

Figure 2.5B presents the cost-effectiveness acceptability curve (CEAC) of the

Table A.11: Scenario analysis results.

	Base	Aging population scenario		Partialuptake scenario	
Prevalence in 2025					
CIT	161,740	189,303	(17.04%)	-	
Oral	199,228	229,286	(15.09%)	198,602	(-0.31%)
Annual cost of CLL management ($)					
2014 CIT	948,972,374	955,479,699	(0.69%)	-	
2016 CIT	994,163,585	1,017,491,537	(2.35%)	-	
2014 Oral	917,874,069	926,149,191	(0.90%)	917,874,069	(0.00%)
2016 Oral	1,657,380,404	1,686,708,467	(1.77%)	1,634,837,050	(-1.36%)
2018 Oral	3,383,614,544	3,503,477,197	(3.54%)	3,161,858,594	(-6.55%)
2020 Oral	4,238,284,439	4,468,901,709	(5.44%)	4,066,862,129	(-4.04%)
2025 Oral	5,130,373,316	5,759,684,042	(12.27%)	5,080,468,797	(-0.97%)
Cumulative cost 2011-2025 ($)					
CIT	15,218,203,943	16,204,937,328	(6.48%)	-	
Oral	44,395,963,354	47,328,146,113	(6.60%)	43,175,278,807	(-2.75%)
Lifetime treatment cost ($)					
CIT: 2016 cohort	115,762	115,754	(-0.01%)	-	
Oral: 2016 cohort	604,071	603,962	(-0.02%)	525,350	(-13.03%)
Oral: 2020 cohort	601,294	602,865	(0.26%)	602,138	(0.14%)
Lifetime out-of-pocket cost ($)					
CIT: 2016 cohort	650	651	(0.30%)	-	
Oral: 2016 cohort	56,560	56,641	(0.14%)	48,421	(-14.39%)
Oral: 2020 cohort	55,332	55,471	(0.25%)	55,406	(0.13%)
Cost-effectiveness (total discounted values for 2011-2025)					
Cost ($): CIT	5,908,502,555	6,342,645,316	(7.35%)	-	
LY: CIT	2,043,997	2,154,196	(5.39%)	-	
QALY: CIT	1,743,135	1,838,279	(5.46%)	-	
Cost ($): Oral	26,075,583,444	28,185,211,686	(8.09%)	24,728,649,769	(-5.17%)
LY: Oral	2,192,509	2,308,518	(5.29%)	2,190,070	(-0.11%)
QALY: Oral	1,849,712	1,949,548	(5.40%)	1,847,635	(-0.11%)
ICER ($/QALY)	189,226	196,305	(3.74%)	180,098	(-4.82%)
ICER ($/LY)	135,795	141,539	(4.23%)	128,840	(-5.12%)

two treatment strategies considered. It clearly shows that oral targeted therapies is deemed cost-effective with a very low probability even at a willingness-to-pay as high as $150,000/QALY, which validates the robustness of the base case ICER estimate and our findings that oral targeted therapies is deemed not cost-effective.

APPENDIX B

APPENDIX FOR CHAPTER 3

B.1 POMDP Formulation

B.1.1 Continuous state DP

In the following, we present the continuous state dynamic program formulation for the liver cancer surveillance problem, which will be shown to be equivalent to a special case of a POMDP model later.

- $\mathscr{S} = \mathscr{S}_d \cup \mathscr{S}_e$: Core state space, including the disease states $\mathscr{S}_d := \{(f,h) : f \in \mathscr{F}, h \in \mathscr{H}\}$ (where $\mathscr{F} := \{0,1,2\}$ and $\mathscr{H} := \{0,1,2,3\}$) and absorbing states $\mathscr{S}_e := \{\mathsf{PTX}, \mathsf{DLD}, \mathsf{DOC}\}$ (the same definitions as in our original manuscript).

- $\mathscr{B} = \{[b(s)] : \sum_{s\in\mathscr{S}} b(s) = 1; 0 \le b(s) \le 1, \forall s \in \mathscr{S}\}$: State distribution over the core states $\mathscr{S}$. For example,

$$\mathbf{b} = (\underbrace{0.1,\ 0.05,\ 0.05,\ 0}_{\text{F3-subpopulation}},\ \underbrace{0.3,\ 0.05,\ 0.05,\ 0}_{\text{CC-subpopulation}},\ \underbrace{0.2,\ 0.1,\ 0.1,\ 0}_{\text{DC-subpopulation}},\quad \underbrace{0,\ 0,\ 0}_{\text{absorbing states}}\quad)$$

  representing the distribution probability of state $(0,0)$, $(0,1)$, $(0,2)$, $(0,3)$, $(1,0)$, $\cdots$, $(2,2)$, $(2,3)$, PTX, DLD, and DOC.

- $\mathscr{A} = \{(a_0, a_1, a_2) : a_f \in \{\mathsf{E}, \mathsf{W}\}, \forall f \in \mathscr{F}\}$: Action space. Because the screening actions can be different across f-subpopulations and that we require them to be the same within a given f-subpopulation, a single action (i.e., either screen or not) is not sufficient to define the decisions for the entire population. Instead, a proper action $\mathbf{a} \in \mathscr{A}$ needs to be a triplet where the each element represents the action in the corresponding f-subpopulation. For example, $\mathbf{a} = (\mathsf{E}, \mathsf{W}, \mathsf{W})$ means screening action (E) in F3-subpopulation ($f = 0$), and no screening (W) in CC and DC-subpopulations ($f = 1, 2$).

- $P^{(a_0,a_1,a_2)}(s'|s)$: Core state transition probability. For $s = (f, h)$, the state transition probability is determined by the action a_f in triplet $\mathbf{a} = (a_0, a_1, a_2)$. In particular, if $a_f = \mathsf{W}$, the transition probability follows the natural progression probability; if $a_f = \mathsf{E}$, after the screening test, some patients are diagnosed with cancer and initiate treatment (i.e., $s' = \mathsf{PTX}$), and the others follow the natural progression. For example, for F3-subpopulation ($f = 0$), the connection between the transition probabilities with actions $a_0 = \mathsf{E}$ and W is formalized as follows:

$$P^{(\mathsf{E},a_1,a_2)}(s'|(f=0,h)) = \begin{cases} sens_s sens_d, \text{ if } h \neq 0 \text{ and } s' = \mathsf{PTX}, \\ (1 - sens_s sens_d) \underbrace{P^{(\mathsf{W},a_1,a_2)}(s'|(f=0,h))}_{\text{Natural progression probability}}, \\ \text{otherwise.} \end{cases}$$

- $r_k(s, \mathbf{a})$: Expected net benefit reward accumulated in period k when the state is s and the action is $\mathbf{a}$.

- $r_N(s)$: Expected terminal net benefit reward at the end of planning horizon for the state s.

- $T[\mathbf{b}, \mathbf{a}]$: Updated state distribution, given the past distribution $\mathbf{b}$ and the action $\mathbf{a}$. Let $T[\mathbf{b}, \mathbf{a}](s')$ denote the component for state s' in the updated state distribution $T[\mathbf{b}, \mathbf{a}]$. It is obvious that

$$T[\mathbf{b}, \mathbf{a}](s') = \sum_{s \in \mathscr{S}} b(s) P^{\mathbf{a}}(s'|s). \tag{B.1}$$

Let $V_k^*(\mathbf{b})$ present the maximum net benefit that the population can attain in period k and onwards given the state distribution $\mathbf{b}$ in period k. Then the optimality equation

is:

$$V_k^*(\mathbf{b}) = \max_{\mathbf{a}\in\mathscr{A}} \left\{ \sum_{s\in\mathscr{S}} b(s) r_k(s,\mathbf{a}) + V_{k+1}^*(T[\mathbf{b},\mathbf{a}]) \right\}, \quad \text{for } k = 1, \cdots, N-1; \tag{B.2}$$

$$V_N^*(\mathbf{b}) = \sum_{s\in\mathscr{S}} b(s) r_N(s). \tag{B.3}$$

A special case POMDP model

The above continuous state DP model is indeed a special case of POMDP models. The updated state distribution $T[\mathbf{b},\mathbf{a}]$ in (B.1), which only depends on the state transition probability and the previous distribution, can be viewed as a Bayesian update of the belief state in a standard POMDP model with the observation space $\Theta = \mathscr{S}$ and a non-informative observation matrix $K^{\mathbf{a}}$, i.e., $K^{\mathbf{a}}(o|s) = \frac{1}{|\mathscr{S}|}$ for all $o \in \Theta, s \in \mathscr{S}, \mathbf{a} \in \mathscr{A}$.

Then we define the belief state updates $\tau[\mathbf{b},\mathbf{a},o]$ given current belief state $\mathbf{b}$, action $\mathbf{a} \in \mathscr{A}$ and observations $o \in \Theta$. Let $\tau[\mathbf{b},\mathbf{a},o](s')$ be the component for state s' in the updated belief $\tau[\mathbf{b},\mathbf{a},o]$, i.e., $\tau[\mathbf{b},\mathbf{a},o](s') = \mathbb{P}(s'|\mathbf{b},\mathbf{a},o)$.

$$\tau(\mathbf{b},\mathbf{a},o)[s'] = \frac{K^{\mathbf{a}}(o|s')\sum_s b(s)P^{\mathbf{a}}(s'|s)}{\sum_j K^{\mathbf{a}}(o|j)\sum_s b(s)P^{\mathbf{a}}(j|s)} = \frac{\frac{1}{|\mathscr{S}|}\sum_s b(s)P^{\mathbf{a}}(s'|s)}{\frac{1}{|\mathscr{S}|}\sum_j\sum_s b(s)P^{\mathbf{a}}(j|s)} = \sum_{s\in\mathscr{S}} b(s)P^{\mathbf{a}}(s'|s),$$

which is indeed reduced to the state distribution update $T[\mathbf{b},\mathbf{a}]$ in (B.1).

Next, we check the equivalence of the optimality equations of the POMDP model and the Equations (B.2)-(B.3). In the POMDP model, we use $U_k^*(\mathbf{b})$ to present the maximum net benefit for period k and onwards given the belief state $\mathbf{b}$. The

optimality equations for the POMDP are as follows:

$$
\begin{aligned}
U_k^*(\mathbf{b}) =& \max_{\mathbf{a}\in\mathscr{A}} \left\{ \sum_{s\in\mathscr{S}} b(s) r_k(s,\mathbf{a}) + \sum_{o\in\Theta} \mathbb{P}(o|\mathbf{b},\mathbf{a}) U_{k+1}^*(\tau[\mathbf{b},\mathbf{a},o]) \right\} \\
=& \max_{\mathbf{a}\in\mathscr{A}} \left\{ \sum_{s\mathscr{S}} b(s) r_k(s,\mathbf{a}) + \sum_{o\in\Theta} \sum_{s'\in\mathscr{S}} K^{\mathbf{a}}(o|s') \sum_{s\in\mathscr{S}} b(s) P^{\mathbf{a}}(s'|s) U_{k+1}^*(\tau[\mathbf{b},\mathbf{a},o]) \right\} \\
=& \max_{\mathbf{a}\in\mathscr{A}} \left\{ \sum_{s\mathscr{S}} b(s) r_k(s,\mathbf{a}) + \frac{1}{|\mathscr{S}|} \sum_{o\in\Theta} U_{k+1}^*(\tau[\mathbf{b},\mathbf{a},o]) \right\} \\
=& \max_{\mathbf{a}\in\mathscr{A}} \left\{ \sum_{s\mathscr{S}} b(s) r_k(s,\mathbf{a}) + U_{k+1}^*(\tau[\mathbf{b},\mathbf{a},o]) \right\}, \quad k = 1,\cdots,N-1, \qquad \text{(B.4)} \\
U_N^*(\mathbf{b}) =& \sum_{s\in\mathscr{S}} b(s) r_N(s), \qquad \text{(B.5)}
\end{aligned}
$$

where Equation (B.4) holds for arbitrary observation $o \in \Theta$ since $\tau[\mathbf{b},\mathbf{a},o]$ is independent of the choice of observation o in Equation (B.1).

Therefore, the continuous state DP formulation (B.2)-(B.3) is reduced to a POMDP formulation (B.4)-(B.5), implying that we could apply the existing POMDP algorithms for solving the continuous state DP model.

B.1.2 POMDP formulation for M-switch policies

To formulate M-switch policies, we need to impose additional constraints to the above POMDP model. Since such constraints define the correlations of decisions across different decision periods, they will break down the principle of optimality (see more detailed explanations in our responses to comment #5), which is essential for solving a DP in a recursive manner. An alternative is to modify the POMDP formulation by augmenting the state space with the screening interval and switch information and modifying the action space accordingly, such that the Markov structure and the principle of optimality will preserve. Obviously, this modification comes at a further computational cost, so the resulting formulation is hopeless to solve computationally. However, we still present it to show that technically it is possible to formulate this problem as a POMDP. The corresponding formulation is presented as follows.

- **State space.** An augmented state $\bar{s} = (\mathbf{b}, \mathbf{d}, \mathbf{g}, \mathbf{w}) \in \bar{\mathscr{S}}$ consisting of belief state $\mathbf{b}$ and "tracker states" $(\mathbf{d}, \mathbf{g}, \mathbf{w})$. In particular,
  - $b(s), s \in \mathscr{S}$: Distribution over core states $\mathscr{S}$.
  - $d_f \in \mathscr{D}, f \in \mathscr{F}$: Current screening interval in f-subpopulation, where $\mathscr{D}$ is the set of possible screening intervals. For example, $\mathscr{D} = \{3M, 6M, 12M, 18M\}$ representing screening every 3-month, 6-month, 12-month, and 18-month.
  - $g_f \in \mathscr{N}, f \in \mathscr{F}$: Number of periods since the last switch of intervals in f-subpopulation. The initial value is 0 at period $k = 1$.
  - $w_f \in \{0, 1, \cdots, M\}, f \in \mathscr{F}$: Number of remaining interval switches in f-subpopulation. The initial value is M (= the predefined maximum number of interval switches for each subpopulation) at period $k = 1$.
- **Action space.** The action space is modified accordingly as follows. Actions are no longer "whether screen or not" decisions; instead, the actions are defined to assign screening intervals: $\bar{\mathbf{a}} \in \bar{\mathscr{A}}_{\bar{s}} = \{(\bar{a}_0, \bar{a}_1, \bar{a}_2) : \bar{a}_f \in \bar{\mathscr{A}}_{\bar{s}}^f, f \in \mathscr{F}\}$, where

$$\bar{\mathscr{A}}_{\bar{s}}^f = \begin{cases} \mathscr{D} & \text{if } w_f \geq 1, \text{ and } g_f = nd_f \text{ for some } n = 0, 1, \cdots, \\ \{d_f\} & \text{if } w_f = 0, \text{ and } g_f = nd_f \text{ for some } n = 0, 1, \cdots, \\ \{\mathsf{NA}\} & \text{if } g_f \neq nd_f \text{ for any } n = 0, 1, \cdots, \end{cases}$$

  where $g_f = nd_f$ for some nonnegative integer n indicates the period where the current screening interval is completed and one can determine the next screening interval (either same or different, subject to the number of remaining possible switches w_f), and the action NA is a dummy action for any other period where the screening interval is not allowed to change.
- **State transitions.**
  - Transitions of state distribution $\mathbf{b}$: The "whether screen or not" decisions (i.e., $\mathbf{a} = (a_f)_{f \in \mathscr{F}}$) for each f-subpopulation are now determined by state

g_f and d_f:

$$a_f = \begin{cases} \mathsf{E} & \text{if } g_f = nd_f \text{ for some } n = 0, 1, \cdots, \\ \mathsf{W} & \text{otherwise.} \end{cases} \tag{B.6}$$

The state distribution $\mathbf{b}$ is then updated following the Equation (B.1) with action $\mathbf{a}$.

– Transitions of d_f, g_f, and w_f. Let d'_f, g'_f and w'_f represent the states in the next period, which are determined as follows:

1. If action $\bar{a}_f \neq \mathsf{NA}$, i.e., a decision of screening interval is made,

$$d'_f = \bar{a}_f,$$

(a) If $\bar{a}_f \neq d_f$, i.e., a different screening interval is assigned,

$g'_f = 1$, (reset the number of periods since the last interval switch)

$w'_f = w_f - 1$. (count down the remaining number of switches by 1)

(b) if $\bar{a}_f = d_f$,

$$g'_f = g_f + 1,$$

$$w'_f = w_f.$$

2. If action $\bar{a}_f = \mathsf{NA}$, i.e., no change in screening interval is made,

$$d'_f = d_f,$$

$$g'_f = g_f + 1,$$

$$w'_f = w_f$$

- **Reward functions.** The intermediate net benefit reward $\bar{r}(\bar{s}, \bar{a})$ accrued in each period can be defined as follows:

$$\bar{r}(\bar{s} = (\mathbf{b}, \mathbf{d}, \mathbf{g}, \mathbf{w}), \bar{\mathbf{a}}) = \sum_{s \in \mathscr{S}} b(s) r(s, \mathbf{a}),$$

where reward $r(s, \mathbf{a})$ is defined in Section B.1.1, and action $\mathbf{a}$ is determined by Equation (B.6).

B.1.3 Computational results of POMDP formulations

We implemented the POMDP model for fully dynamic policy (without M-switch structures) following the above formulation and attempted to solve it numerically, using the *pomdp-solve* package, a widely used computational package for POMDPs (see `http://www.pomdp.org/`). We remark that considering M-switch policies would require further expansion of the space space and hence computational time for the fully dynamic policy would serve as a lower bound for solving the M-switch policies.

We tested the model on the Linux operating system with an Intel Xeon 2.30GHz CPU and 62Gb RAM. The algorithm was able to only proceed for 3 periods with fast growing number of alpha vectors, and exploded afterwards (see the results in the table below). Note that the POMDP formulation for M-switch policies consists of an even larger state space with a more complex structure (a mixture of discrete "tracker" states and continuous belief states in the augmented state space, see Section B.1.2). Thus, solving the POMDP model for identifying the optimal M-switch would be computationally even more demanding.

Iteration	# alpha vectors	Total running time (sec)
1	6	0
2	6	0.02
3	107	1.43
4	-	>7000

B.2 Reduction of HS-M-G to HS-M-NI Formulation

In the following, we will show that the formulation for nested screening interval in our original manuscript can be derived from the above-introduced generalized formulation. In particular, Constraints (3.10)-(3.11) in our manuscript can be derived from Constraints (3.19)-(3.22) in this section given the nested interval structure.

First we have the following observations for the nested structure of screening intervals:

- $\mathscr{J}_1 \supset \mathscr{J}_2 \supset \cdots \supset \mathscr{J}_{|\mathscr{D}|} = \emptyset$.
- If screening interval d_i is selected, all screening periods $\mathscr{J}_{i'}$ for interval $d_{i'}, \forall i' \geq i$ are "automatically" selected.

Due to the "overlapping" nature of nested intervals, we now implicitly allow "multiple" screening intervals to be selected together (e.g., $\mathscr{J}_3$, $\mathscr{J}_4$, and onwards), instead of "one-and-only-one" is allowed as defined in Constraint (3.21). In particular, we define $\phi^f_{l,i}$ to be 1 if screening periods $\mathscr{J}_i$ are selected for cycle l in f-subpopulation, and 0 otherwise. By this definition, we relpace constraints (3.19)-(3.20) with

$$y_{(l-1)C+j}((f,h),\mathsf{E}) = \phi^f_{l,i}, \quad \forall j \in \mathscr{J}_i, i = \{1,\cdots,|\mathscr{D}|\}, f \in \mathscr{F}, h \in \mathscr{H}, l \in \mathscr{L}, \text{ (B.7)}$$

and the constraint (3.21) becomes trivially true and can be removed since we at least have screenings in $\mathscr{J}_{|\mathscr{D}|} = \emptyset$ (i.e., no screening).

Note that there are repeatedly defined equality constraints due to the overlapping $\mathscr{J}$'s. That is, Constraint (B.7) for $\mathscr{J}_{i+1}$ has been defined in the constraints for $\mathscr{J}_i$. Thus, it is sufficient to define the constraints for $j \in \mathscr{J}_i \backslash \mathscr{J}_{i+1}$ as follows:

$$y_{(l-1)C+j}((f,h),\mathsf{E}) = \phi^f_{l,i}, \quad \forall j \in \mathscr{J}_i \backslash \mathscr{J}_{i+1}, i = \{1,\cdots,|\mathscr{D}|-1\}, f \in \mathscr{F}, h \in \mathscr{H}, l \in \mathscr{L}. \text{ (B.8)}$$

Following the previous observations that all screening periods $\mathscr{J}_{i'}$ for interval $d_{i'}, \forall i' \geq i$ are "automatically" selected when $\mathscr{J}_i$ is selected, we have additional constraints:

$$\phi^f_{l,i} \leq \phi^f_{l,i+1}, \ \forall i = \{1,\cdots,|\mathscr{D}|-1\}, f \in \mathscr{F}, l \in \mathscr{L}, \text{ (B.9)}$$

By the change of variables using Constraint (B.7), Constraint (B.9) is as the same as the follows:

$$y_{(l-1)C+j}((f,h),\mathsf{E}) \leq y_{(l-1)C+j'}((f,h),\mathsf{E}),$$
$$\forall j \in \mathscr{J}_i, j' \in \mathscr{J}_{i+1}, i = \{1,\cdots,|\mathscr{D}|-1\}, f \in \mathscr{F}, h \in \mathscr{H}, l \in \mathscr{L}, \text{ (B.10)}$$

which can be further simplified by removing the repeated definitions on $\mathscr{J}_i \cap \mathscr{J}_{i+1}$:

$$y_{(l-1)C+j}((f,h),\mathsf{E}) \leq y_{(l-1)C+j'}((f,h),\mathsf{E}),$$
$$\forall j \in \mathscr{J}_i \backslash \mathscr{J}_{i+1}, j' \in \mathscr{J}_{i+1}, i = \{1,\cdots,|\mathscr{D}|-1\}, f \in \mathscr{F}, h \in \mathscr{H}, l \in \mathscr{L}, \tag{B.11}$$

To sum up, given the nested structure of screening intervals, the general formulation can be simplified as Constraints (B.8) and (B.11). Given that variable $\phi_{l,i}^f$ only appear in Constraint (B.8), we can simply present Constraint (B.8) as

$$y_{(l-1)C+j}((f,h),\mathsf{E}) = y_{(l-1)C+j'}((f,h),\mathsf{E}),$$
$$\forall j, j' \in \mathscr{J}_i \backslash \mathscr{J}_{i+1}, i = \{1,\cdots,|\mathscr{D}|-1\}, f \in \mathscr{F}, h \in \mathscr{H}, l \in \mathscr{L}. \tag{B.12}$$

Finally, the general formulation is simplified to Constraints (B.8) and (B.12), which are identical with the Constraints (3.10)-(3.11).

B.3 Appendix for Analytical Results

B.3.1 Preliminaries

We first establish several relations of quantities previously defined in Section 3.4. For any given policy π, $f \in \mathscr{F}$ and $h \in \mathscr{H}$, value functions have the following backward induction form:

$$V_k^\pi(f,h) = r_k\big((f,h),\pi_k(f)\big)$$
$$+ \gamma \sum_{f' \in \mathscr{F}, h' \in \overline{\mathscr{H}}} P_k\big((f',h')|(f,h),\pi_k(f)\big) V_{k+1}^\pi(f',h'),\ k = 1,\cdots,N-1,$$
$$V_N^\pi(f,h) = r_N(f,h),$$

and for the augmented state $h = 4$, $V_k^\pi(f, h=4) = 0$, $\forall f \in \mathscr{F}, k \in \mathscr{N}$, as no reward is accumulated in absorbing states.

By the construction of policies π^1 and π^2 in Section 3.4 (identical everywhere except for the specific f-subpopulation in period t), we obtain the following relation

between $V_k^{\pi^1}$ and $V_k^{\pi^2}$ in periods $k \geq t$:

$$V_t^{\pi^2}(f,h) = \begin{cases} V_t^{\pi^1}(f,h) - \kappa_0, & \text{if } h = 0, \\ \zeta R_t(f,h) + (1-\zeta)V_t^{\pi^1}(f,h) - \kappa_1, & \text{if } h \in \mathscr{H}^+, \end{cases}$$

$$V_k^{\pi^2}(\tilde{f},\tilde{h}) = V_k^{\pi^1}(\tilde{f},\tilde{h}), \quad \text{if } k = t,\ \tilde{f} \neq f; \text{ or } k \geq t+1,\ \forall \tilde{f} \in \mathscr{F},\ \tilde{h} \in \mathscr{H} \tag{B.13}$$

which indicates that for cancer-free patients, *Screen* action does not change the outcomes except incurring additional test cost κ_0; and for cancer patients, following the *Screen* action with test cost κ_1, $\zeta \times 100\%$ of these patients are diagnosed and receive immediate treatment, while others are missed by the tests and continue the natural progression until the next period.

Let $q'_k(f,h)$, $c'_k(f,h)$, and $\rho_k(f,h)$ respectively represent the expected QALYs, costs, and survival probabilities under treatment for a detected cancer state (f,h), $f \in \mathscr{F},\ h \in \mathscr{H}^+$ in period k. Then, given the terminal QALY $q'_N(f,h)$ and cost $c'_N(f,h)$, the lump-sum treatment NB is calculated recursively as follows:

$$R_k(f,h) = \left[\lambda q'_k(f,h) - c'_k(f,h)\right] + \gamma \rho_k(f,h) R_{k+1}(f,h), \quad \forall f \in \mathscr{F},\ h \in \mathscr{H}^+,\ k = 1, \cdots, N-1, \tag{B.14}$$

$$R_N(f,h) = \lambda q'_N(f,h) - c'_N(f,h), \quad \forall f \in \mathscr{F},\ h \in \mathscr{H}^+, \text{ and}$$

$$R_k(f, h=4) = 0,\ f \in \mathscr{F}, k = 1, \cdots, N. \tag{B.15}$$

If the treatment is delayed by one period (i.e., a cancer patient does not have any intervention at period k and receives treatment at period $k+1$), the lump-sum NB $W_k(f,h)$ in state (f,h), $h \in \mathscr{H}^+$ at period k is calculated as:

$$W_k(f,h) = \left[\lambda q_k((f,h),\mathsf{W}) - c_k((f,h),\mathsf{W})\right] + \gamma \sum_{f',h'} P_k((f',h')|(f,h),\mathsf{W}) R_{k+1}(f',h'), \quad \forall f \in \mathscr{F},\ h \in \mathscr{H}^+,\ k = 1, \cdots, N-1,$$

We let $Q_k^R(f,h)$ and $Q_k^W(f,h)$ denote the QALY components of $R_k(f,h)$ and $W_k(f,h)$, and $C_k^R(f,h)$ and $C_k^W(f,h)$ denote the cost components, respectively. Then,

the QALY and cost components of immediate treatment benefit $(R_k(f,h) - W_k(f,h))$ are defined respectively as follows:

$$\Delta Q_k(f,h) = Q_k^R(f,h) - Q_k^W(f,h) \quad \text{and} \tag{B.16}$$

$$\Delta C_k(f,h) = C_k^R(f,h) - C_k^W(f,h), \quad \forall f \in \mathscr{F},\ h \in \mathscr{H}^+,\ k = 1, \cdots, N-1. \tag{B.17}$$

B.3.2 Proofs of theorems

Lemma B.1 (Fenwick et al. (2001)) *If a policy π has the maximum NB, then π is the most cost-effective policy.*

Proof of Lemma B.1. We prove this by contradiction. Let π have the maximum NB. Suppose there exists a policy π' which is more cost-effective than π. By definition, $E(\pi') > E(\pi)$ and $\frac{C(\pi')-C(\pi)}{E(\pi')-E(\pi)} < \lambda$. Then, $\lambda E(\pi) - C(\pi) < \lambda E(\pi') - C(\pi')$, which is a contradiction. □

Throughout the remainder of this appendix, we omit the superscript π in V_{k+1}^{π} for simplicity, except where its inclusion is necessary to avoid ambiguity. Before presenting the proof of Theorem 3.1, we first introduce a lemma where we characterize several relationships between $V_k^{\pi}(f,h)$, $R_k(f,h)$, and $W_k(f,h)$.

Lemma B.2 *The followings hold for $\forall f \in \mathscr{F}$, $h \geq 1$, and period $k = 1, 2, \cdots, N-1$:*

1. *$V_k^{\pi}(f,h) \leq R_k(f,h)$ for any policy π,*
2. *$V_k^{\pi}(f,h) \leq W_k(f,h)$ if $\pi_k(f) = \mathsf{W}$.*

Proof of Lemma B.2.

Condition 1. By induction.

(1) It holds for $n = N$, because $V_N^{\pi}(f,4) = R_N(f,4) = 0$, and $V_N^{\pi}(f,h) = r_N(f,h) \leq R_N(f,h)$ for $f \in \mathscr{F}$, $h \in \mathscr{H}^+$ by Assumption 3.2.

(2) Suppose the statement holds for $n = k+1$, i.e., for any policy π, $V_{k+1}^{\pi}(f,h) \leq R_{k+1}(f,h)$ for all f and $h \geq 1$. Then, for $n = k$,

a) For policy π^1 where $\pi_k^1(f) = \mathsf{W}$, $\forall h \in \mathscr{H}^+$:

$$V_k^{\pi^1}(f,h) = r_k((f,h),\mathsf{W}) + \gamma \sum_{f'} F_k(f'|f) \sum_{h'} H_k(h'|h,f) V_{k+1}^{\pi^1}(f',h')$$

$$\leq r_k((f,h),\mathsf{W}) + \gamma \sum_{f'} F_k(f'|f) \sum_{h'} H_k(h'|h,f) R_{k+1}(f',h') \tag{B.18}$$

$$= W_k(f,h) \tag{B.19}$$

$$\leq R_k(f,h), \tag{B.20}$$

where Inequality (B.18) follows from the induction hypothesis and (B.20) follows from Assumption 3.3. We remark that (B.20) also holds for absorbing state $h = 4$ because $V_k^{\pi}(f, h = 4) = R_k(f, h = 4) = 0$ for all k.

(b) For policy π^2 where $\pi_k^2(f) = \mathsf{E}$, we define policy π^1 as the policy identical to π^2 except that $\pi_k^1(f) = \mathsf{W}$. Then by Equation (B.13), we have:

$$V_k^{\pi^2}(f,h) = \zeta R_k(f,h) + (1-\zeta) V_k^{\pi^1}(f,h) - \kappa_1 \leq R_k(f,h) - \kappa_1 \leq R_k(f,h),$$

where the first inequality follows from (B.20), which completes the proof.

Condition 2. This condition directly follows from (B.19), where we have shown $V_k^{\pi}(f,h) \leq W_k(f,h)$ if $\pi_k(f) = \mathsf{W}$ for all f and $h \geq 1$ at any period k. □

Proof of Proposition 3.1. For the entire population, the difference of the expected NB between the two policies π^1 and π^2 is $\sum_{\tilde{f},\tilde{h}} b_1(\tilde{f},\tilde{h}) \left(V_1^{\pi^2}(\tilde{f},\tilde{h}) - V_1^{\pi^1}(\tilde{f},\tilde{h}) \right)$. Since the actions in the two policies are different only in period t for the f-subpopulation, this difference is reduced to the marginal surveillance benefit $\Delta_t(f)$ for the f-subpopulation in period t, i.e., $\sum_{\tilde{f},\tilde{h}} b_1(\tilde{f},\tilde{h}) \left(V_1^{\pi^2}(\tilde{f},\tilde{h}) - V_1^{\pi^1}(\tilde{f},\tilde{h}) \right) = \sum_{h \in \mathscr{H}} b_t(f,h)[V_t^{\pi^2}(f,h) -$

$V_t^{\pi^1}(f,h)]$. Then,

$$\begin{aligned}
\Delta_t(f) =& \sum_{h\in\mathscr{H}} b_t(f,h)\left[V_t^{\pi^2}(f,h) - V_t^{\pi^1}(f,h)\right] \\
=& b_t(f,0)\left[\left(V_t^{\pi^1}(f,0)-\kappa_0\right) - V_t^{\pi^1}(f,0)\right] \\
&+ \sum_{h\in\mathscr{H}^+} b_t(f,h)\left[\left(\zeta R_t(f,h) + (1-\zeta)V_t^{\pi^1}(f,h) - \kappa_1\right) - V_t^{\pi^1}(f,h)\right] \\
=& -b_t(f,0)\kappa_0 + \sum_{h\in\mathscr{H}^+} b_t(f,h)\left[\zeta\left(R_t(f,h) - V_t^{\pi^1}(f,h)\right) - \kappa_1\right] \qquad \text{(B.21)} \\
\geq& -b_t(f,0)\kappa_0 + \sum_{h\in\mathscr{H}^+} b_t(f,h)\left[\zeta\left(R_t(f,h) - W_t(f,h)\right) - \kappa_1\right],
\end{aligned}$$

where the inequality follows from Lemma B.2. By the NB definitions of $R_k(f,h)$ and $W_k(f,h)$,

$$\begin{aligned}
\Delta_t(f) \geq& -b_t(f,0)\kappa_0 \\
&+ \sum_{h\in\mathscr{H}^+} b_t(f,h)\left[\zeta\left(\left[\lambda Q_t^R(f,h) - C_t^R(f,h)\right] - \left[\lambda Q_t^W(f,h) - C_t^W(f,h)\right]\right) - \kappa_1\right] \\
=& \lambda\zeta \sum_{h\in\mathscr{H}^+} b_t(f,h)\Delta Q_t(f,h) - \left[b_t(f,0)\kappa_0 + \left(\sum_{h\in\mathscr{H}^+} b_t(f,h)\right)\kappa_1\right] \\
&- \zeta \sum_{h\in\mathscr{H}^+} b_t(f,h)\Delta C_t(f,h) \qquad \text{(B.22)} \\
\geq& 0, \qquad \text{(B.23)}
\end{aligned}$$

where (B.22) follows from the definitions of $\Delta Q_t(f,h)$ and $\Delta C_t(f,h)$ given in (B.16) and (B.17), respectively, and (B.23) follows from Condition (3.27), which completes the proof. □

We now introduce two lemmas which are used in proving other statements.

Lemma B.3 *Let $y = [y_s]$ represent a column vector and y_s represent the s-th element of the vector. If y_s is decreasing in s, and transition matrix $\mathbf{P}$ is IFR, then $(\mathbf{P}y)_s$ is decreasing in s.*

Proof of Lemma B.3. The proof follows from Lemma 4.7.2 in (175). □

Lemma B.4 *Let* $\mathbf{P}$ *and* $\mathbf{Q}$ *be two* $n \times n$ *transition matrices. If* $\mathbf{P} \preceq \mathbf{Q}$ *and vector* $y = [y_s] \in \mathbb{R}^n$ *is decreasing in* s*, then* $\mathbf{P}y \geq \mathbf{Q}y$ *component-wise, i.e.,* $(\mathbf{P}y)_s \geq (\mathbf{Q}y)_s$, $\forall s = 1, \cdots, n$.

Proof of Lemma B.4. The result immediately follows from Definition 3.2 and Lemma 4.7.2 in (175). □

Proposition B.1 (Monotonicity) *For any given policy* π*, the value function* $V_k^\pi(f, h)$ *is decreasing in both* $f \in \mathscr{F}$ *and* $h \in \overline{\mathscr{H}}$.

Proof of Proposition B.1. First we show that for any given policy π, $V_k^\pi(f, 0) \geq R_k(f, 1)$ for any f-subpopulation and period $k \in \mathscr{N}$. By induction,

(1) For $n = N$, $V_N(f, 0) = r_N(f, 0) \geq R_N(f, 1)$ by Assumption 3.2.

(2) Assume $V_n^\pi(f, 0) \geq R_n(f, 1)$ holds for $n = k+1$, and let $\pi_k(f)$ represent the action taken in period k for the f-subpopulation. Then, for $h = 0$ and $n = k$:

$$
\begin{aligned}
V_k^\pi(f,0) =& r_k((f,0),\pi_k(f)) + \gamma \sum_{f'} \sum_{h'} P_k((f',h')|(f,0),\mathsf{W}) V_{k+1}(f',h') \\
=& \Big(\lambda q_k((f,0),\mathsf{W}) - c_k((f,0),\mathsf{W}) - \mathbf{1}\{\pi_k(f) = \mathsf{E}\}\kappa_0\Big) \\
& + \gamma \sum_{f'} \sum_{h'} P_k((f',h')|(f,0),\mathsf{W}) V_{k+1}(f',h') \qquad \text{(B.24)} \\
\geq& \Big(\lambda q_k((f,0),\mathsf{W}) - c_k((f,0),\mathsf{W}) - \mathbf{1}\{\pi_k(f) = \mathsf{E}\}\kappa_0\Big) \\
& + \gamma P_k((f,0)|(f,0),\mathsf{W}) V_{k+1}(f,0) \qquad \text{(B.25)} \\
\geq& (\lambda q_k'(f,1) - c_k'(f,1)) + \gamma \rho_k(f,1) R_{k+1}(f,1) \qquad \text{(B.26)} \\
=& R_k(f,1), \qquad \text{(B.27)}
\end{aligned}
$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Equation (B.24) follows from the definition of intermediate NB $r_k(s, a)$, (B.25) follows by dropping other non-negative terms in the summation, (B.26) follows from Assumption 3.6 and the inductive hypothesis, and (B.27) follows from the definition of $R_k(f, h)$ (B.14). Thus, $V_k^\pi(f, 0) \geq R_k(f, 1)$ holds true for any fibrosis state f and all periods.

Next, we prove the monotonicity results by induction.

(1) For $n = N$, this is true for any given $f \in \mathscr{F}$ and $h \in \mathscr{H}$ by Assumption 3.2. Since $V_N(f,h) = r_N(f,h) \geq 0,\ \forall h \in \mathscr{H}$ and $V_N(f, h = 4) = 0$ for absorbing state $h = 4$, it is also true for $f \in \mathscr{F}$ and $h \in \overline{\mathscr{H}}$.

(2) Suppose the statement holds for $n = k + 1$. That is, for any policy π, $V^{\pi}_{k+1}(f,h)$ is decreasing in $f \in \mathscr{F}$ and $h \in \overline{\mathscr{H}}$. Then, for $n = k$,

Case 1: The action in fibrosis state f in period k is *Wait*, i.e., the policy taken is π^1 where $\pi^1_k(f) = \mathsf{W}$. Recall the definition of value function $V^{\pi^1}_k(f,h)$:

$$V^{\pi^1}_k(f,h) = r_k((f,h),\mathsf{W}) + \gamma \sum_{f'} F_k(f'|f) \sum_{h'} H_k(h'|h,f) V_{k+1}(f',h').$$

1. Fix fibrosis state f. We have that $\sum_{h'} H_k(h'|h,f)V_{k+1}(f',h')$ is decreasing in h for each f' by Assumption 3.4, inductive hypothesis, and Lemma B.3. Then, the linear combination $\sum_{f'} F_k(f'|f) \sum_{h'} H_k(h'|h,f)V_{k+1}(f',h')$ with non-negative coefficients $F_k(f'|f)$ is also decreasing in h. Since $r_k((f,h),\mathsf{W})$ is decreasing in h by Assumption 3.2, $V^{\pi^1}_k$ is decreasing in h for any given fibrosis state f.

2. Fix tumor state h. For $\tilde{f} \geq f$, by the inductive hypothesis, Assumption 3.4, and Lemma B.4, we have $\sum_{h'} H_k(h'|h,f)V_{k+1}(f',h') \geq \sum_{h'} H_k(h'|h,\tilde{f})V_{k+1}(f',h')$ for all $f' \in \mathscr{F}$. Then,

$$\begin{aligned}
&\sum_{f'} F_k(f'|f) \sum_{h'} H_k(h'|h,f) V_{k+1}(f',h') \\
\geq &\sum_{f'} F_k(f'|f) \sum_{h'} H_k(h'|h,\tilde{f}) V_{k+1}(f',h') \\
\geq &\sum_{f'} F_k(f'|\tilde{f}) \sum_{h'} H_k(h'|h,\tilde{f}) V_{k+1}(f',h'),
\end{aligned}$$

where the last inequality follows from Assumption 3.4 and Lemma B.3. Since $r_k((f,h),\mathsf{W}) \geq r_k((\tilde{f},h),\mathsf{W})$ by Assumption 3.2 , we have for $f \leq \tilde{f}$,

$$V^{\pi^1}_k(f,h) \geq r_k((\tilde{f},h),\mathsf{W}) + \gamma \sum_{f'} F_k(f'|\tilde{f}) \sum_{h'} H_k(h'|h,\tilde{f}) V_{k+1}(f',h') = V^{\pi^1}_k(\tilde{f},h).$$

Thus, $V^{\pi^1}_k(f,h)$ is decreasing in f for any given tumor state h.

Case 2: The action for fibrosis state f in period k is *Screen*, i.e., the policy taken is π^2 where $\pi^2_k(f) = \mathsf{E}$, and let π^1 be the policy identical to π^2 except that $\pi^1_k(f) = \mathsf{W}$.

1. For a fixed f:

   a) When $h = 0$, by the definition of value function $V_k^{\pi_2}$ given in (B.13),

$$\begin{aligned} V_k^{\pi^2}(f,0) =& V_k^{\pi^1}(f,0) - \kappa_0 \\ \geq& \zeta R_k(f,1) + (1-\zeta) V_k^{\pi^1}(f,0) - \kappa_0 \\ \geq& \zeta R_k(f,1) + (1-\zeta) V_k^{\pi^1}(f,1) - \kappa_1 \\ =& V_k^{\pi^2}(f,1) \end{aligned} \tag{B.28}$$

   where the first inequality follows from (B.27), the second inequality follows because $V_k^{\pi^1}(f,0) \geq V_k^{\pi^1}(f,1)$ by Case 1(1) and the fact that $\kappa_0 \leq \kappa_1$ by Assumption 3.5, and (B.28) follows again from the definition of $V_k^{\pi^2}$ in (B.13).

   b) When $h \in \mathscr{H}^+$, by (B.13), the value function is equal to:

$$V_k^{\pi^2}(f,h) = \zeta R_k(f,h) + (1-\zeta) V_k^{\pi^1}(f,h) - \kappa_1.$$

   Since both $R_k(f,h)$ and $V_k^{\pi^1}(f,h)$ are decreasing in h by Assumption 3.2 and Case 1, we have $V_k^{\pi^2}(f,h)$ is decreasing in $h \in \mathscr{H}^+$. To show that $V_k^{\pi^2}$ is decreasing in $h \in \overline{\mathscr{H}} = \mathscr{H} \cup \{4\}$, it is sufficient to show that $V_k^{\pi^2}(f,h) \geq 0$ for any $h \in \mathscr{H}^+$ because $V_k^{\pi^2}(f, h=4) = 0$. Below we show that $V_k^{\pi^2}(f,h) \geq 0$ for any $h \in \mathscr{H}^+$.

$$\begin{aligned} V_k^{\pi^2}(f,h) =& V_k^{\pi^1}(f,h) + \zeta(R_k(f,h) - V_k^{\pi^1}(f,h)) - \kappa_1 \\ \geq& V_k^{\pi^1}(f,h) + \zeta(R_k(f,h) - W_k(f,h)) - \kappa_1 \geq 0, \end{aligned}$$

   which follows from Lemma B.2 and Assumption 3.3. Therefore, $V_k^{\pi^2}(f,h)$ is decreasing in $h \in \overline{\mathscr{H}}$ for a fixed fibrosis state f.

2. For a fixed h, $V_k^{\pi^2}(f,h)$ is decreasing in f because both $R_k(f,h)$ and $V_k^{\pi^1}(f,h)$ in Equation (B.13) are decreasing in f by Assumption 3.2 and Case 1(2), which completes the proof.

□

Proof of Proposition 3.2. By induction.

(1) As terminal rewards are the same, i.e., for any $f \in \mathscr{F}$, $V_N^1(f,h) = V_N^2(f,h) = r_N(f,h)$ for $h \in \mathscr{H}$ and $V_N^1(f,h=4) = V_N^2(f,h=4) = 0$ for absorbing state $h=4$, the assertion holds for $n = N$.

(2) Assume the assertion holds for $n = k+1$. That is, $V_{k+1}^1(f,h) \geq V_{k+1}^2(f,h), \forall f \in \mathscr{F}$ and $h \in \overline{\mathscr{H}}$. Then, the following inequalities hold true for any given f and h,

$$\sum_{h'} H_k^1(h'|f,h)V_{k+1}^1(f',h') \geq \sum_{h'} H_k^2(h'|f,h)V_{k+1}^1(f',h') \geq \sum_{h'} H_k^2(h'|f,h)V_{k+1}^2(f',h'). \tag{B.29}$$

where the first inequality follows from Lemma B.4 with the given condition $\mathbf{H}_k^1(f) \succeq \mathbf{H}_k^2(f)$ and the monotonicity of value function from Proposition B.1; the second inequality follows from the induction hypothesis that $V_{k+1}^1(f,h) \geq V_{k+1}^2(f,h), \forall f \in \mathscr{F}$ and $h \in \overline{\mathscr{H}}$.

Since f and h are given, by Lemma B.3, for $i = 1, 2$, $\sum_{h'} H_k^i(h'|f,h)V_{k+1}^i(f',h')$ is decreasing in f'. Then, by applying a similar argument as in (B.29), we have

$$\sum_{f'} F_k^1(f'|f)\Big(\sum_{h'} H_k^1(h'|f,h)V_{k+1}^1(f',h')\Big) \geq \sum_{f'} F_k^2(f'|f)\Big(\sum_{h'} H_k^2(h'|f,h)V_{k+1}^2(f',h')\Big). \tag{B.30}$$

Now we show the statement holds for $n = k$. As before, let π^1 and π^2 be two identical policies with the same actions, except that the action is *Wait* in π^1 and *Screen* in π^2 for state f in period k. Recall that any policy can be presented as either π^1 and π^2 because for any policy, the action taken for f-subpopulation in a particular period t is either W or E. Let $V_k^{\pi,i}$ represent the value function for policy π associated with disease progressions $\{\mathbf{F}_k^i, \mathbf{H}_k^i(f)\}$.

1. For policy π^1 where the action taken in state f in period k is *Wait* (i.e., $\pi_k^1(f) = \mathsf{W}$):

$$V_k^{\pi^1,i}(f,h) = r_k((f,h),\mathsf{W}) + \gamma \sum_{f'} F_k^i(f'|f) \sum_{h'} H_k^i(h'|h,f) V_{k+1}^i(f',h'),\ i = 1,2. \tag{B.31}$$

Then $V_k^{\pi^1,1}(f,h) \geq V_k^{\pi^1,2}(f,h), \forall f,h$, which follows from (B.30).

2. For policy π^2 where the action taken in state f in period k is *Screen* (i.e., $\pi_k^2(f) = \mathsf{E}$), we let π^1 be the policy identical to π^2 except that $\pi_k^1(f) = \mathsf{W}$:

$$\begin{aligned} V_k^{\pi^2,1}(f,0) - V_k^{\pi^2,2}(f,0) =& \big(V_k^{\pi^1,1}(f,0) - \kappa_0\big) - \big(V_k^{\pi^1,2}(f,0) - \kappa_0\big) \\ =& \big(V_k^{\pi^1,1}(f,0) - V_k^{\pi^1,2}(f,0)\big) \geq 0, \end{aligned}$$

and

$$\begin{aligned} V_k^{\pi^2,1}(f,h) - V_k^{\pi^2,2}(f,h) =& \big(\zeta R_k(f,h) + (1-\zeta) V_k^{\pi^1,1}(f,h) - \kappa_1\big) \\ & - \big(\zeta R_k(f,h) + (1-\zeta) V_k^{\pi^1,2}(f,h) - \kappa_1\big) \\ =& (1-\zeta)\big(V_k^{\pi^1,1}(f,h) - V_k^{\pi^1,2}(f,h)\big) \geq 0, \quad \forall h \geq 1, \end{aligned}$$

where both results follow from the first case shown above, which completes the proof. □

Proof of Theorem 3.1. Recall that by Equation (B.21):

$$\Delta_t^i(f) = -b_t(f,0)\kappa_0 + \sum_{h \in \mathscr{H}^+} b_t(f,h) \left[\zeta\big(R_t(f,h) - V_t^{\pi^1,i}(f,h)\big) - \kappa_1\right], \quad \text{for } i = 1,2.$$

From Proposition 3.2, we know that $V_t^{\pi^1,1}(f,h) \geq V_t^{\pi^1,2}(f,h)$ for all $f \in \mathscr{F}$ and $h \in \overline{\mathscr{H}}$. Therefore, it follows that $R_t(f,h) - V_t^{\pi^1,1}(f,h) \leq R_t(f,h) - V_t^{\pi^1,2}(f,h)$ for all $f \in \mathscr{F}$ and $h \in \mathscr{H}^+$. Thus, $\Delta_t^1(f) \leq \Delta_t^2(f)$ for all f, which completes the proof. □

Before proving Theorem 3.2, we present two more lemmas which are used in proving this theorem.

Lemma B.5 *Let $\{b_k^1(f,h)\}$ and $\{b_k^2(f,h)\}$ represent two state distributions corresponding to two HCC progression matrices $\{\mathbf{H}_k^1(f)\}$ and $\{\mathbf{H}_k^2(f)\}$ (with same fibrosis progression matrices $\{\mathbf{F}_k\}$), respectively, where $\mathbf{H}_k^1(f) \succeq \mathbf{H}_k^2(f)$ for all f and period k. Suppose the initial distributions are the same, $b_1^1 = b_1^2$. Then, for any given policy $\pi \in \Pi$, $f \in \mathscr{F}$, and $k \in \mathscr{N}$, we have $\sum_{h \geq l} b_k^1(f,h) \geq \sum_{h \geq l} b_k^2(f,h)$, $\forall l \in \overline{\mathscr{H}}$.*

Proof of Lemma B.5. By induction.

(1) It is trivial for $k = 1$.

(2) Suppose $\sum_{h \geq l} b_n^1(f,h) \geq \sum_{h \geq l} b_n^2(f,h)$ holds for $n = k$ for all $f \in \mathscr{F}$ and $l \in \overline{\mathscr{H}}$. Let $\mathscr{F}^{\mathsf{W}}$ and $\mathscr{F}^{\mathsf{E}}$ denote the fibrosis states which have *Wait* and *Screen* actions at period k, respectively. Then, transition probabilities $P_k((f,h)|(f',h'),\mathsf{W}) = H_k(h|h',f')F_k(f|f'), \forall f' \in \mathscr{F}^{\mathsf{W}}$ and $P_k((f,h)|(f',h'),\mathsf{E}) = G_k(h|h',f')F_k(f|f'), \forall f' \in \mathscr{F}^{\mathsf{E}}$, where transition matrix $\mathbf{G}_k(f)$ is defined based on $\mathbf{H}_k(f)$ as follows:

$$\mathbf{G}_k(f) = \begin{bmatrix} H_k(0|0,f) & \cdots & H_k(3|0,f) & H_k(4|0,f) \\ (1-\zeta)H_k(0|1,f) & \cdots & (1-\zeta)H_k(3|1,f) & (1-\zeta)H_k(4|1,f)+\zeta \\ (1-\zeta)H_k(0|2,f) & \cdots & (1-\zeta)H_k(3|2,f) & (1-\zeta)H_k(4|2,f)+\zeta \\ (1-\zeta)H_k(0|3,f) & \cdots & (1-\zeta)H_k(3|3,f) & (1-\zeta)H_k(4|3,f)+\zeta \\ & & & 1 \end{bmatrix} \tag{B.32}$$

It is easy to verify that $\mathbf{G}_k(f)$ is also IFR. In addition, if $\mathbf{H}_k^1(f) \succeq \mathbf{H}_k^2(f)$, then $\mathbf{G}_k^1(f) \succeq \mathbf{G}_k^2(f)$ for any period k, by Definition 3.2 and Equation (B.32).

Then, when $n = k+1$, for any $l \in \overline{\mathscr{H}}$,

$$\begin{aligned} \sum_{h \geq l} b_{k+1}(f,h) &= \sum_{h \geq l} \sum_{f',h'} b_k(f',h') P_k\big((f,h)|(f',h'),\pi_k(f')\big) \\ &= \sum_{f' \in \mathscr{F}^{\mathsf{W}}} F_k(f|f') \sum_{h'} b_k(f',h') \sum_{h \geq l} H_k(h|h',f') \\ &\quad + \sum_{f' \in \mathscr{F}^{\mathsf{E}}} F_k(f|f') \sum_{h'} b_k(f',h') \sum_{h \geq l} G_k(h|h',f') \end{aligned} \tag{B.33}$$

1) For $f' \in \mathscr{F}^{\mathsf{W}}$, note that

$$\sum_{h'} b_k^1(f',h') \sum_{h \geq l} H_k^1(h|h',f') \geq \sum_{h'} b_k^1(f',h') \sum_{h \geq l} H_k^2(h|h',f') \geq \sum_{h'} b_k^2(f',h') \sum_{h \geq l} H_k^2(h|h',f'), \tag{B.34}$$

where the first inequality follows from the fact that $\mathbf{H}_k^1(f) \succeq \mathbf{H}_k^2(f)$, and the second inequality follows from the inductive hypothesis and the fact that $\sum_{h \geq l} H_k^2(h|h',f')$ is increasing in h' due to the IFR property (Assumption 3.4), and Lemma 4.7.2 in (175). Since $\mathbf{F}_k$ is the same for $\{b_k^1(f,h)\}$ and $\{b_k^2(f,h)\}$, we have

$$\sum_{f' \in \mathscr{F}^{\mathsf{W}}} F_k(f|f') \sum_{h'} b_k^1(f',h') \sum_{h \geq l} H_k^1(h|h',f') \geq \sum_{f' \in \mathscr{F}^{\mathsf{W}}} F_k(f|f') \sum_{h'} b_k^2(f',h') \sum_{h \geq l} H_k^2(h|h',f').$$

2) For $f' \in \mathscr{F}^{\mathsf{E}}$, we can apply the same argument in Equation (B.34) since $\mathbf{G}_k^1$ and $\mathbf{G}_k^2$ have the same properties as $\mathbf{H}_k^1$ and $\mathbf{H}_k^2$. Therefore, we have

$$\sum_{f' \in \mathscr{F}^{\mathsf{E}}} F_k(f|f') \sum_{h'} b_k^1(f',h') \sum_{h \geq l} G_k^1(h|h',f') \geq \sum_{f' \in \mathscr{F}^{\mathsf{E}}} F_k(f|f') \sum_{h'} b_k^2(f',h') \sum_{h \geq l} G_k^2(h|h',f'),$$

and thus

$$\sum_{h \geq l} b_{k+1}^1(f,h) \geq \sum_{h \geq l} b_{k+1}^2(f,h),$$

which completes the proof. □

Next, in the following lemma, we show that $R_k(f,h) - V_k^{\pi^1}(f,h)$ is decreasing in h.

Lemma B.6 *Suppose policy π^1 has action* Wait *for fibrosis state f in period t, i.e., $\pi_t^1(f) = \mathsf{W}$. For a fixed fibrosis state f and any tumor state $h = 1, 2$, if*

$$R_t(f,h) - R_t(f,h+1) \geq \Big[r_t((f,h),\mathsf{W}) - r_t((f,h+1),\mathsf{W}) \Big] + \gamma H_t(h+1|h,f) R_{t+1}(f,h+1), \tag{B.35}$$

then,

$$R_t(f,h) - V_t^{\pi^1}(f,h) \geq R_t(f,h+1) - V_t^{\pi^1}(f,h+1).$$

Proof of Lemma B.6. Fix fibrosis state f,

$$
\begin{aligned}
& V_t^{\pi}(f,h) - V_t^{\pi}(f,h+1) \\
=& r_t((f,h),\mathsf{W}) + \gamma \sum_{f' \geq f} F_t(f'|f) \sum_{h' \geq h} H_t(h'|h,f) V_{t+1}(f',h') \\
& - \Big(r_t((f,h+1),\mathsf{W}) + \gamma \sum_{f' \geq f} F_t(f'|f) \sum_{h' \geq h+1} H_t(h'|h+1,f) V_{t+1}(f',h') \Big) \qquad \text{(B.36)} \\
=& \underbrace{\big[r_t((f,h),\mathsf{W}) - r_t((f,h+1),\mathsf{W}) \big]}_{:=\tilde{r}_t(f,h)} + \gamma \sum_{f' \geq f} F_t(f'|f) \Big[H_t(h+1|h,f) V_{t+1}(f',h+1) \\
& + \sum_{h' \geq h+1} \underbrace{\big(H_t(h'|h,f) - H_t(h'|h+1,f) \big)}_{\leq 0} V_{t+1}(f',h') \Big] \qquad \text{(B.37)} \\
\leq& \tilde{r}_t(f,h) + \gamma \sum_{f' \geq f} F_t(f'|f) H_t(h+1|h,f) V_{t+1}(f',h+1) \\
\leq& \tilde{r}_t(f,h) + \gamma \sum_{f' \geq f} F_t(f'|f) H_t(h+1|h,f) R_{t+1}(f',h+1) \qquad \text{(B.38)} \\
\leq& \tilde{r}_t(f,h) + \gamma \left(\sum_{f' \geq f} F_t(f'|f) \right) H_t(h+1|h,f) R_{t+1}(f,h+1) \qquad \text{(B.39)} \\
=& \tilde{r}_t(f,h) + \gamma H_t(h+1|h,f) R_{t+1}(f,h+1) \\
\leq& R_t(f,h) - R_t(f,h+1), \qquad \text{(B.40)}
\end{aligned}
$$

where (B.36) and $H_t(h'|h,f) - H_t(h'|h+1,f) \leq 0$ in (B.37) follow from Assumption 3.4, (B.38) follows from the fact that $V_{t+1}(f,h) \leq R_{t+1}(f,h)$ by Lemma B.2, (B.39) follows from that $R_t(f,h)$ is decreasing in f by Assumption 3.2, and (B.40) immediately follows from Condition (B.35). □

Proof of Theorem 3.2. Policy π^1 represents a policy with the action *Wait* for f-subpopulation in period t, i.e., $\pi_t^1(f) = \mathsf{W}$. We define $U_t(f,h) := R_t(f,h) - V_t^{\pi^1}(f,h)$ for $h \in \mathscr{H}^+$ to represent the treatment benefit. Let $\Delta R_t(f,h)$, $\Delta U_t(f,h)$, and $\Delta V_t^{\pi^1}(f,h)$ represent the change in $R_t(f,h)$, $U_t(f,h)$, and value function $V_t^{\pi^1}(f,h)$, respectively, as the risk of developing HCC decreases and treatment NB increases.

Then, by definition given in (3.26):

$$\begin{aligned}\Delta_t^1(f) =& - b_t^1(f,0)\kappa_0 + \sum_{h\in\mathscr{H}^+} b_t^1(f,h)\Big(\zeta\Big(R_t(f,h) - V_t^{\pi_1}\Big) - \kappa_1\Big),\\ \Delta_t^2(f) =& - b_t^2(f,0)\kappa_0 + \sum_{h\in\mathscr{H}^+} b_t^2(f,h)\Big(\zeta\Big[\Big(R_t(f,h) + \Delta R_t(f,h)\Big)\\ & - \Big(V_t^{\pi_1}(f,h) + \Delta V_t^{\pi_1}(f,h)\Big)\Big] - \kappa_1\Big).\end{aligned}$$

Also, by the definition of $U_t(f,h)$:

$$\begin{aligned}\Delta_t^2(f) - \Delta_t^1(f) =& - b_t^2(f,0)\kappa_0 + \sum_{h\in\mathscr{H}^+} b_t^2(f,h)\left[\zeta\Big(U_t(f,h) + \Delta U_t(f,h)\Big) - \kappa_1\right]\\ & - \left[-b_t^1(f,0)\kappa_0 + \sum_{h\in\mathscr{H}^+} b_t^1(f,h)\Big(\zeta U_t(f,h) - \kappa_1\Big)\right]\\ =& - (b_t^2(f,0) - b_t^1(f,0))\kappa_0\\ & + \sum_{h\in\mathscr{H}^+}\left[b_t^2(f,h)\zeta\Delta U_t(f,h) - \Big(b_t^2(f,h) - b_t^1(f,h)\Big)\Big(\zeta U_t(f,h) - \kappa_1\Big)\right].\end{aligned}$$

Let $\{\mathbf{H}_k^1(f)\}$ represent the HCC progression probability matrix for the base case, and $\{\mathbf{H}_k^2(f)\}$ for the case with reduced HCC risk and improved treatment outcome. Clearly, $\mathbf{H}_k^1(f) \succeq \mathbf{H}_k^2(f)$ for all f and period k. By Lemma B.5, we have $\sum_{h\geq l} b_t^1(f,h) \geq \sum_{h\geq l} b_t^2(f,h),\ l \in \overline{\mathscr{H}}$. In addition, we have that $\zeta U_t(f,h) - \kappa_1$ is decreasing in $h \in \mathscr{H}^+$ by Condition (3.28) and Lemma B.6, and is nonnegative by Assumption 3.3. Thus, by Lemma 4.7.2 in (175),

$$\begin{aligned}&\Big[b_t^2(f,1), b_t^2(f,2), b_t^2(f,3), b_t^2(f,4)\Big]\Big[\zeta U_t(f,1) - \kappa_1, \zeta U_t(f,2) - \kappa_1, \zeta U_t(f,3) - \kappa_1, 0\Big]'\\ \geq&\Big[b_t^1(f,1), b_t^1(f,2), b_t^1(f,3), b_t^1(f,4)\Big]\Big[\zeta U_t(f,1) - \kappa_1, \zeta U_t(f,2) - \kappa_1, \zeta U_t(f,3) - \kappa_1, 0\Big]'\end{aligned} \tag{B.41}$$

$$\iff \sum_{h\in\mathscr{H}^+} b_t^2(f,h)\big(\zeta U_t(f,h) - \kappa_1\big) \geq \sum_{h\in\mathscr{H}^+} b_t^1(f,h)\big(\zeta U_t(f,h) - \kappa_1\big),$$

which implies that $\Delta_t^2(f) - \Delta_t^1(f) \leq -(b_t^2(f,0) - b_t^1(f,0))\kappa_0 + \sum_{h\in\mathscr{H}^+} b_t^2(f,h)\zeta\Delta U_t(f,h)$.

Note that $\Delta V_t^\pi(f,h) \geq 0$ for $h \in \mathscr{H}^+$. This is because (1) when treatment NB $R_k(f,h)$ does not change, value function $V_k^\pi(f,h)$ increases as HCC risk reduces

by Proposition 3.2 and the fact that $\mathbf{H}_k^1(f) \succeq \mathbf{H}_k^2(f)$; and (2) when treatment NB $R_k(f,h)$ increases, value function $V_k^\pi(f,h)$ will further increase if policy π has more than one *Screen* actions after period t, which can be checked by the recursive form of value functions and (B.13). Thus, $\Delta U_t(f,h) = \Delta R_t(f,h) - \Delta V_t^\pi(f,h) \leq \Delta R_t(f,h)$, which implies

$$\begin{aligned}
\Delta_t^2(f) - \Delta_t^1(f) \leq & - [b_t^2(f,0) - b_t^1(f,0)]\kappa_0 + \sum_{h \in \mathscr{H}^+} b_t^2(f,h)\zeta \Delta R_t(f,h) \\
\leq & - [b_t^2(f,0) - b_t^1(f,0)]\kappa_0 + \left(\sum_{h \in \mathscr{H}^+} b_t^2(f,h) \right) \zeta \max_{h \in \mathscr{H}^+} \{\Delta R_t(f,h)\} \\
\leq & 0,
\end{aligned}$$

where the last inequality follows from Condition (3.29) in Theorem 3.2, which completes the proof.

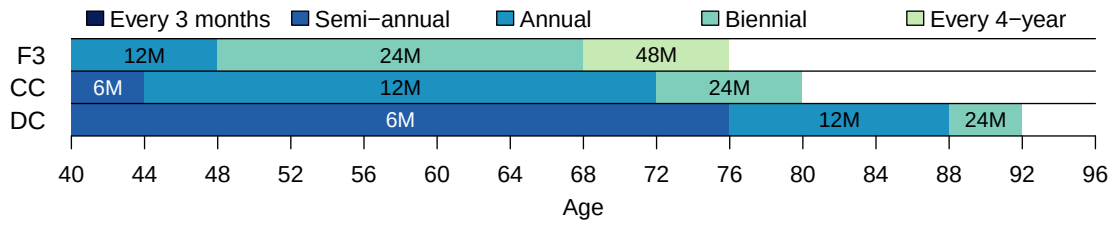
□

B.4 Extended Sensitivity Analysis Results

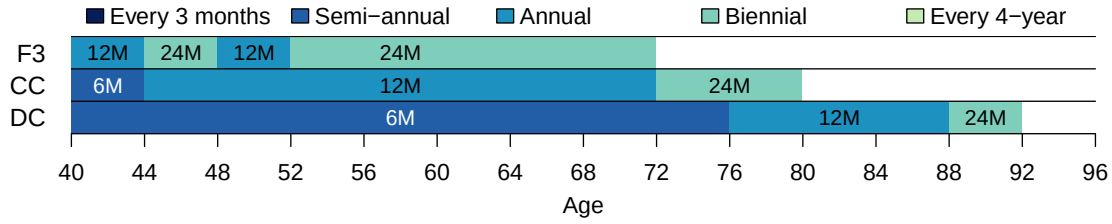
B.4.1 Policies with more interval switches

To examine the effect of the possible number of switches (M), on the optimal M-switch policies, we solve the HS-M problem with up to 5 switches of surveillance intervals. Figure B.1 presents these optimal M-switch policies and Table B.1 summarizes the performance of corresponding policies.

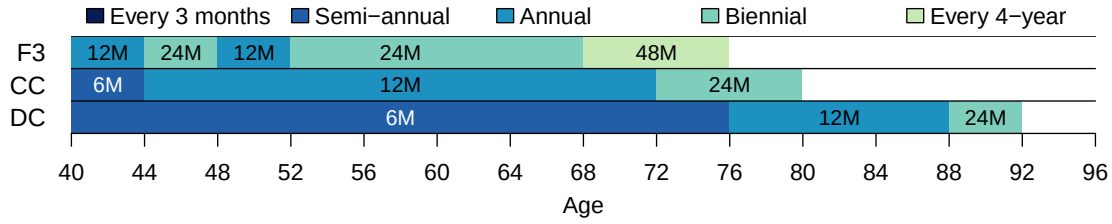
We observe that the increasing number of maximum possible switches does not change the policy for CC and DC subpopulations at all, and has limited changes for F3 subpopulation. Also, the allowed additional switches do not translate into significant improvement in policy performance. In Table B.1, we observe that the total NB has only marginal improvement with the increasing number of switches, which implies that more than one switch does not substantially improve the cost-effectiveness and health outcomes of M-switch policies.



(a) 2-switch policy



(b) 3-switch policy



(c) 4-switch/5-switch policy

Figure B.1: Sensitivity analysis for the number of switches.

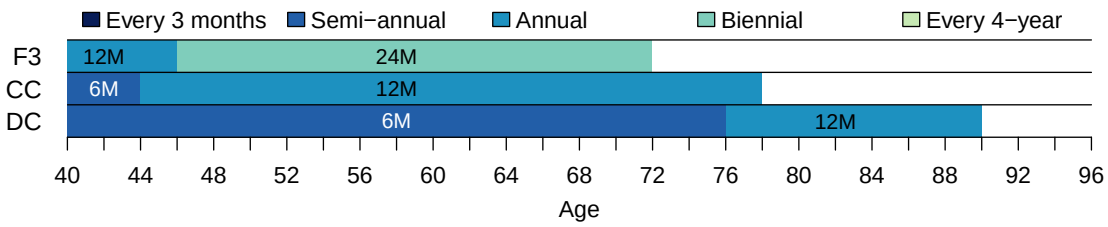
Table B.1: M-switch policies with up to 5 switches.

Surveillance policy	QALY	LY	Cost($)	# Tests	# per 1,000 patients			
					# HCC detected			DLD
					Small	Medium	Large	
1-switch	347424.7	7.791	10.133	42139	11.7	125	39	7	516
2-switch	347426.1	7.791	10.133	42125	11.7	125	39	7	516
3-switch	347427.1	7.790	10.132	42091	11.5	125	39	7	516
4-switch	347427.1	7.790	10.132	42091	11.5	125	39	7	516
5-switch	347427.1	7.790	10.132	42091	11.5	125	39	7	516

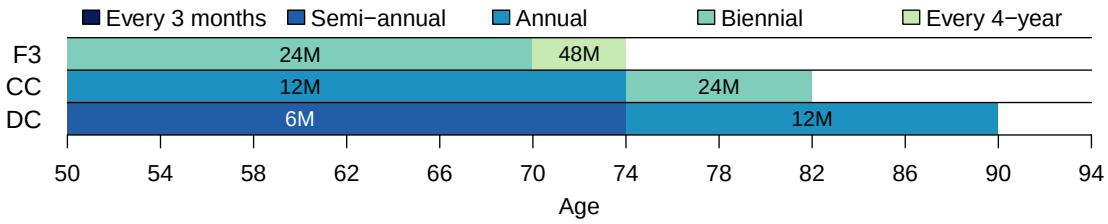
B.4.2 Scenario analysis

We next examine the optimal surveillance policies under different settings of switching cycle, initial population age, and WTP value. For each scenario, we present the results for 1-switch policy, as it well represents the class of dynamic policies with a relatively

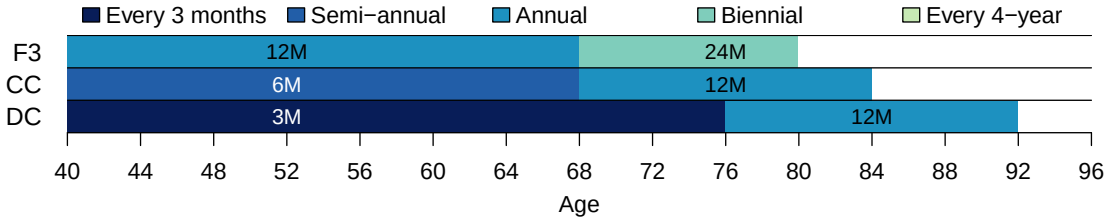
simple and practical structure.



(a) 1-switch policy with 2 year switching cycle



(b) 1-switch policy for 50-year-old population



(c) 1-switch policy for WTP=$100,000/QALY

Figure B.2: Sensitivity analysis for switching cycle length, initial age, and WTP.

We find that policy results are not very sensitive to the choice of the switching cycle or the initial population age. When we use a longer switching cycle which allows longer surveillance intervals, the model does not utilize any interval longer than 4 years and leads to the same policy as in the base case results; when we reduce the switching cycle to 2 years which enables quicker switches to different intervals, the optimal 1-switch policy has only minimal changes in surveillance intervals compared with the base case results (Figure B.2a). The optimal 1-switch policy for a population with initial age 50 is also similar to the base case results (Figure B.2b).

When WTP value is higher, the optimal 1-switch policy recommends more frequent surveillance, as expected. For example, Figure B.2c presents the optimal 1-switch policy with the WTP value of $100,000/QALY. This result also reflects the tendency of utilizing more frequent surveillance in future, as higher WTP values are more likely to be accepted by policy makers in future (166).

B.4.3 Probabilistic sensitivity analysis

In addition to the PSA results presented in Section 3.5.4, we also assessed the cost-effectiveness acceptability curves (CEACs) by comparing stratified and 1-switch policies with the annual surveillance policy, the most cost-effective routine policy in F3 and cirrhotic patients (see Table 3.7). Figure B.3 shows that, compared with annual surveillance policy, 0-switch and 1-switch policies also achieve very high confidence of cost-effectiveness for a wide range of WTPs. Thus, the PSA results demonstrate that our model results are robust against parameter uncertainties, and thus further support our findings.

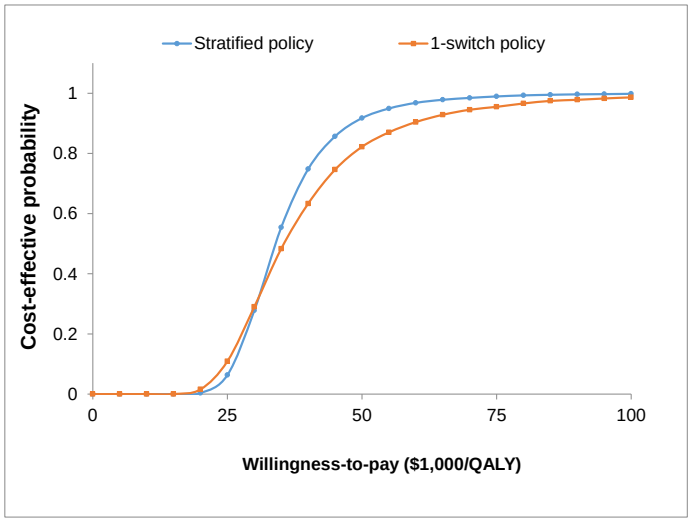


Figure B.3: Cost-effectiveness acceptability curve for optimal stratified policies and 1-switch policy compared with every-12 month surveillance.

B.5 Parameter Estimation

B.5.1 Calibration of HCC risks based on HALT-C trial

Since existing HCC incidence estimates in the literature do not represent the transition probabilities from cancer-free state to the small HCC state in our model, we calibrate such transition probabilities based on a recent clinical study, the HALT-C trial (130). The HALT-C trial was designed to evaluate the long-term outcomes in patients who had chronic hepatitis C infection with advanced fibrosis or compensated cirrhosis and were not cured by previous hepatitis C treatment. Since it is the most recent and the only study that has long-term follow-up for non-cirrhotic patients, the HALT-C trial represents the best clinical data for our model calibration.

The HALT-C trial has a medium follow-up time of 6.5 years, in which 622 (59%) patients are with advanced fibrosis (i.e., F3 stage) and 428 (41%) patients are with cirrhosis at their enrollment with an average age of 50. We use the same initial patient characteristics in our model for the calibration. The primary goal of calibration is to identify reasonable estimates for the transition probabilities such that the percentage of total detected HCC cases from the model will be similar to those from the clinical data. Although the protocol of the HALT-C trial recommends ultrasound surveillance every 12-month during the first 3.5 years and every 6 months during the extended follow-up, only 68.9% patients have consistent screenings at least once per year (201). Thus, we evaluate annual and biennial surveillance policies in our model and use the weighted average of the results from the two policies (68.9% for the annual surveillance) as the model output for the calibration. To search for reasonable estimates of the transition probabilities, we use the incidence estimates from the literature as the initial value, and then increase the value until the model outputs fall in the 95% confidence intervals of the estimates based on clinical data. Table B.2 shows the calibration results for the estimated probabilities of developing small HCC that are selected as the base values (in Table 3.2).

Table B.2: Calibration results for the risk of developing HCC

Performance measure	Estimates in HALT-C % (95% CI)	Model output (%)	
		Weighted	Annual/Biennial
HCC% at 3.5 years (144)	3.40 (2.35, 4.70)	2.36	2.77/1.44
HCC% at median follow-up (6.5 years) (144)	4.80 (3.54, 6.28)	3.59	3.79/3.14
Annual risk of death/transplant (F3) (60)	2.20 (1.20, 3.68)	1.98	1.96/2.02
Annual risk of death/transplant (CC) (60)	5.30 (3.38, 7.87)	6.43	6.38/6.54

HCC, hepatocellular carcinoma; CI, confidence interval; CC, compensated cirrhosis.

B.5.2 Transition probabilities

Transition probabilities in the model are derived from disease progression risks, mortality risks, and test accuracy parameters. To estimate the transition matrix for natural progression $[P_k(s'|s, \mathsf{W})]$, we need to specify the transition matrices $\{\mathbf{F}_k, \mathbf{H}_k(f)\}$ (see Assumption 3.1).

In transition matrix $\mathbf{F}_k = [F_k(f'|f)]$ for fibrosis states, we assume that in each 3-month (i.e., one period in our model), fibrosis stage can either progress to the next advanced stage or remain unchanged. This is because the fibrosis progression is usually slow, which typically takes about one decade to progress to the next advanced stage (217). Estimates of annual risks from F3 to CC and from CC to DC are available in published meta-analysis and systematic reviews (217, 173). We then convert the annual risk into the risk for three months using the formula $r' = 1 - (1 - r)^{T'/T}$ as suggested in (164), where r and r' represent the risk for time horizon T=12 months and T'=3 months, respectively.

Transition matrix $\mathbf{H}_k(f)$ for tumor states is estimated based on the following components:

- $p_1(f)$: Monthly risk of developing small HCC for each fibrosis state f, which is obtained from the model calibration (see Section B.5.1).

- p_2, p_3: Monthly risks of HCC progression from small to medium, and medium

to large HCC, respectively, which are estimated from (51).

- $\widehat{\mathbf{H}}_k(f) = [\widehat{H}_k(h'|h,f)]$: 3-month transition matrix for HCC progression for each fibrosis state f, which is calculated by

$$\widehat{\mathbf{H}}_k(f) = \left(\begin{bmatrix} 1-p_1(f) & p_1(f) & & \\ & 1-p_2 & p_2 & \\ & & 1-p_3 & p_3 \\ & & & 1 \end{bmatrix} \right)^3$$

- $m^l(f,h)$: Mortality from liver disease state (f,h). For the early stage HCC (i.e., $h \leq 2$), we assume the tumor does not lead to the excess mortality. In this case, the excess mortality only depends on the fibrosis state f, which is estimated from the systematic review of clinical studies (54). For the late stage HCC (i.e., $h = 3$), as it is deadly and usually leads to very poor prognosis (with a median survival of 6 months), the excess mortality for large HCC is high for all non-tumor fibrosis state f and is estimated based on the mortality of patients undergoing best supportive care in a retrospective analysis (91).

- m_k^o: Age-dependent mortality from other causes, which are estimated from the US life tables (15).

- $m_k(f,h) := m^l(f,h) + m_k^o$: Total mortality for state (f,h).

We define the transition matrix $\mathbf{H}_k(f)$ as follows:

$$H_k(h'|h,f) = \begin{cases} \widehat{H}_k(h|h,f) - m_k(f,h) & \text{if } h' = h, \\ \widehat{H}_k(h'|h,f) & \text{otherwise,} \end{cases}$$

which satisfies

$$\sum_{h' \in \mathscr{H}} H_k(h'|h,f) + m_k(f,h) = 1, \quad \forall f \in \mathscr{F},\ h \in \mathscr{H},\ k = 1, \cdots, N-1.$$

Then, the natural progression transition probabilities are characterized as follows:

$$P_k(s'|(f,h),\mathsf{W}) = \begin{cases} F_k(f'|f)H_k(h'|h,f) & \text{if } s' = (f',h') \in \mathscr{S}_d, \\ m^l(f,h) & \text{if } s' = \mathsf{DLD}, \\ m_k^o & \text{if } s' = \mathsf{DOC}. \end{cases} \tag{B.42}$$

Accuracy of surveillance and diagnostic tests are estimated from the literature. For the surveillance ultrasound test, we use the sensitivity of 0.75 and specificity of 0.94 following Anderson et al. (13). For the HCC diagnosis, MRI and CT are the commonly used diagnostic imaging tests. Note that sometimes in practice a second imaging test may be needed if the first one is not conclusive. Thus, using the estimate of a single test may underestimate the actual diagnostic accuracy. Considering this, we estimate the diagnostic sensitivity as 0.9, which is in line with the range reported in literature (182, 50).

B.5.3 Lump-sum treatment reward

Since no surveillance decisions are considered after HCC treatment, we assign a lump-sum treatment reward in the absorbing state PTX in our model, which is calculated from separate Markov models. In clinical practice, four main types of treatment are commonly used: liver resection, liver transplant, ablation, and palliative treatment. We estimate treatment distributions based on (13, 70), and expert opinion, which are presented in Table B.3.

- *Resection, ablation, and palliative treatment* are usually delivered soon after the HCC diagnosis. We use a simple 2-state Markov model (Figure B.4) to calculate the total quality-adjusted survival time for each treatment separately. For each treatment, post-treatment mortality is determined by the age-dependent background mortality (15) plus the treatment-specific mortality. Treatment-specific

mortality, derived based on the survival estimates from clinical studies, depend on f and h states. Quality of life (QOL) after each type of treatment is determined by the QOL at treatment initiation at disease state (f, h) (see Section B.5.4) and the *percentage change* in QOL after the corresponding treatment as estimated in (220) (converted to EQ5D utilities based on method in (45)). We consider these treatments as one-time procedures without maintenance therapy for a long-term follow-up, and thus use a fixed cost for each treatment estimated based on a population study (125).

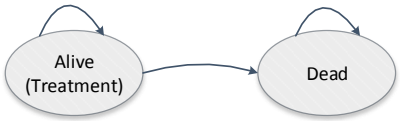


Figure B.4: Two-state Markov model for estimating lump-sum rewards of liver resection, ablation, and palliative treatment.

- *Liver transplant* is different from other treatments in the sense that transplant-eligible patients need to wait for additional time on the waiting list until an available organ is available for the transplant. While waiting, patients may lose their eligibility for transplant as they may progress outside the Milan criteria meanwhile. In that case, they are managed by palliative treatment (see state transitions in Figure B.5). We estimate the probability of transplant on the waiting list in each period based on the average waiting time of transplant recipients from the United Network for Organ Sharing (UNOS) data (168). After transplant, patients have a long-term follow-up with maintenance costs in subsequent years. Estimates for post-transplant survival, QOL, and treatment costs are available from the literature (see Table 3.3).

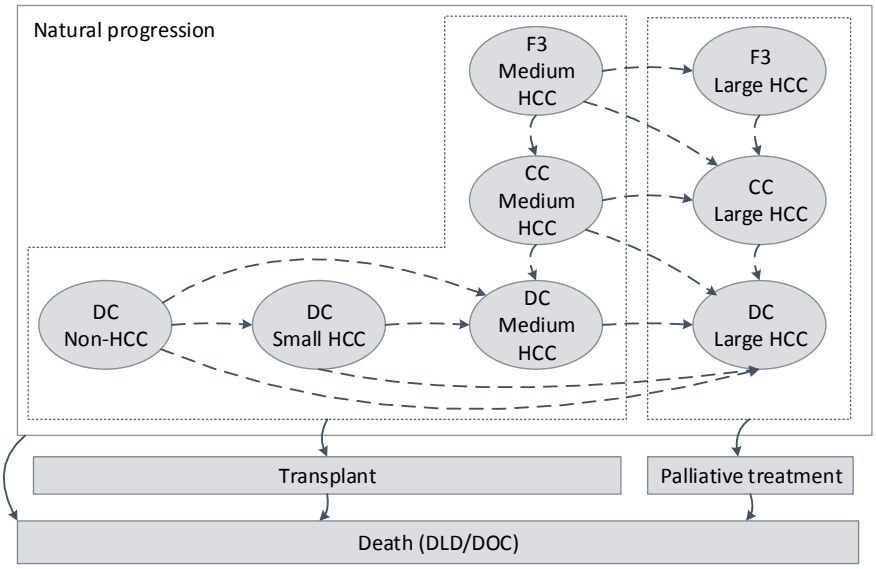


Figure B.5: Markov model for estimating lump-sum rewards of liver transplant.

Table B.3: Treatment distribution

Non-tumor state	Tumor state	RES	OLT	ABL	PAL
F3/CC	small HCC	35%		35%	30%
F3/CC	medium HCC	28%	14%	28%	30%
F3/CC	large HCC	13%		6%	81%
DC	small/medium HCC		28%	42%	30%
DC	large HCC			19%	81%

RES, resection; OLT, liver transplant; ABL, ablation; PAL, palliative treatment; CC, compensated cirrhosis; DC, decompensated cirrhosis.

B.5.4 Intermediate QALY and cost

For health states (f, h), we assume any *undetected* HCC state h does not impair the quality of life as the development and progression of HCC are usually silent with rare clinical symptoms (51). That is, the utility for state (f, h) only depends on the fibrosis state f, denoted as $u(f)$. In addition, we adjust the utility by age to capture

the aging effect (97). That is, the utility in state $s \in \mathscr{S}$ is given by:

$$u_k(s) := \begin{cases} u(f)\eta_k, & \text{if } s = (f,h) \in \mathscr{S}_d,\ \forall f \in \mathscr{F}, h \in \mathscr{H}, \\ 0, & \text{if } s \in \mathscr{S}_e. \end{cases}$$

where η_k represents the age-adjustment factor (97). The intermediate QALYs $q_k(s,a,s')$ from state s to s' with action a is defined by:

$$q_k(s,a,s') := \begin{cases} \frac{1}{2}\left(u_k(s) + u_{k+1}(s')\right) & \text{if } a = \mathsf{W}, \text{ or } a = \mathsf{E} \text{ and } s' \neq \mathsf{PTX}, \\ Q_k^R(s) & \text{if } a = \mathsf{E} \text{ and } s' = \mathsf{PTX}, \\ 0 & \text{otherwise}, \end{cases}$$

where we take the average of $u_k(s)$ and $u_{k+1}(s')$ for the half-cycle correction, and $Q_k^R(s)$ represents the lump-sum QALYs of treatment for state $s \in \mathscr{S}_d$. Similarly, we define the cost in state $s \in \mathscr{S}$ as:

$$c_k(s) := \begin{cases} c(f) & \text{if } s = (f,h) \in \mathscr{S}_d,\ \forall f \in \mathscr{F}, h \in \mathscr{H}, \\ 0 & \text{if } s \in \mathscr{S}_e, \end{cases}$$

where $c(f)$ is the cost associated with the fibrosis state f for the routine care of chronic hepatitis C. Then, we define the intermediate cost from state s to s' with action a as follows:

$$c_k(s, a = \mathsf{W}, s') = \frac{1}{2}\left(c_k(s) + c_{k+1}(s')\right)$$

$$c_k(s, a = \mathsf{E}, s') = \begin{cases} \kappa_0 + \frac{1}{2}\left(c_k(s) + c_{k+1}(s')\right) & \text{if } s = (f,h),\ \forall f \in \mathscr{F},\ h = 0, \text{ and } s' \neq \mathsf{PTX}, \\ \kappa_1 + \frac{1}{2}\left(c_k(s) + c_{k+1}(s')\right) & \text{if } s = (f,h),\ \forall f \in \mathscr{F},\ h \geq 1, \text{ and } s' \neq \mathsf{PTX}, \\ c_s + c_d + C_k^R(s) & \text{if } s = (f,h),\ \forall f \in \mathscr{F},\ h \geq 1, \text{ and } s' = \mathsf{PTX}, \\ 0 & \text{otherwise}, \end{cases}$$

where $C_k^R(s)$ represent the lump-sum cost of treatment for state $s \in \mathscr{S}_d$.

B.5.5 Initial HCC prevalence

The initial state distribution requires estimates of the prevalence of *unobservable* HCC for each f-subpopulation. However, such estimates do not exist in the literature. Therefore, to estimate initial prevalence, we use the prevalence of projected HCC cases from a population-based simulation model (110). These estimates are stratified for each 5-year age groups (see Table B.4). Since the simulation model does not differentiate the HCC cases by the tumor size at diagnosis, we assume the simulation-projected HCC prevalence corresponds to the the distribution of small HCC in our base case result.

B.5.6 Model Validation

In addition to validating the clinical implications of our findings with domain experts, to validate our model parameterization, we have cross-compared our model results with a previously published and well-recognized cost-effectiveness study by Andersson et al. (13). In particular, Anderson et al. assessed the cost-effectiveness of (routine) screening policies with different fixed intervals for 50-year-old compensated cirrhosis (CC) patients. To compare our model results to those in (13), we applied our model to the same population profile and screening policies. Our model produced comparable results to those of Andersson's model (Table B.5), despite existing differences in modeling details.

Table B.4: Initial HCC prevalence

	F3		CC		DC	
Age group	Non-HCC	HCC	Non-HCC	HCC	Non-HCC	HCC
30	1,575 (70.00%)	20 (0.89%)	618 (27.47%)	10 (0.44%)	27 (1.20%)	0 (0.00%)
35	16,434 (37.50%)	246 (0.56%)	24,053 (54.88%)	537 (1.23%)	2,266 (5.17%)	289 (0.66%)
40	28,959 (35.23%)	498 (0.61%)	46,680 (56.80%)	998 (1.21%)	4,456 (5.42%)	598 (0.73%)
45	29,209 (35.10%)	492 (0.59%)	47,224 (56.75%)	1,144 (1.37%)	4,543 (5.46%)	605 (0.73%)
50	55,764 (34.10%)	935 (0.57%)	94,439 (57.75%)	2,102 (1.29%)	9,053 (5.54%)	1,235 (0.76%)
55	46,720 (34.21%)	821 (0.60%)	78,631 (57.57%)	1,834 (1.34%)	7,592 (5.56%)	976 (0.71%)
60	64,715 (33.67%)	1,110 (0.58%)	111,697 (58.11%)	2,672 (1.39%)	10,595 (5.51%)	1,431 (0.74%)
65	45,534 (33.58%)	747 (0.55%)	78,875 (58.16%)	1,797 (1.33%)	7,645 (5.64%)	1,014 (0.75%)
70	39,027 (33.34%)	670 (0.57%)	68,454 (58.47%)	1,574 (1.34%)	6,468 (5.52%)	878 (0.75%)

Table B.5: Model validation result

	QALY		Cost ($)	
	Our model	Andersson(2008)	Our model	Andersson(2008)
No screening	6.13	5.97	29662	26170
Annual screening (ultrasound)	6.39	6.35	36149	34161
Semiannual screening (ultrasound)	6.43	6.45	38238	37272

APPENDIX C

APPENDIX FOR CHAPTER 4

C.1 Proofs of Propositions

Proof of Proposition 4.1. First observe that $\dot{N} = \Lambda - \mu(S+R) - m(I+D) = \Lambda - \mu N - (m-\mu)(I+D) \leq \Lambda - \mu N$ because $m \geq \mu$ by definition. Let $\dot{M} = \Lambda - \mu M$ with $M(0) = M_0 = \Lambda/\mu$. Then $M(t) = \Lambda/\mu$ for $t \geq 0$. By the comparison theorem (Theorem 1.5.2 in (124)), if $N(0) = N_0 \leq M_0$, then $N(t) \leq M(t) = \Lambda/\mu$ for $t \geq 0$.

Next we consider the case $N_0 > \frac{\Lambda}{\mu}$. Note that $N(t) \leq N_0 + \int_0^t \Lambda - \mu N(\tau) d\tau$, where $\lambda - \mu N(\tau) < 0$ when $N(\tau) > \Lambda/\mu$ for $\tau \in [0,t]$. And thus $N(t) < N_0$ holds true for all time before $N(t) = \Lambda/\mu$. If $N(t') = \Lambda/\mu$ at some time t', $N(t) \leq \Lambda/\mu$ for $t \geq t'$ by the previous argument. □

Proof of Proposition 4.2. We first determine the disease-free equilibrium (DFE) by solving

$$\begin{cases} \dot{S} = \dot{I} = \dot{D} = \dot{R} = 0 \\ I = D = 0 \end{cases} \implies \begin{cases} \dot{S} = \Lambda - \mu S = 0 \\ \dot{R} = -\mu R = 0. \end{cases}$$

which leads to the DFE $\mathcal{E} = (S^*, I^*, D^*, R^*) = \left(\frac{\Lambda}{\mu}, 0, 0, 0\right)$.

We employ the next-generation operator method (58) to calculate the reproduction number $\mathscr{R}_c$. For the infected compartments I and D, we define the new infections matrix $\mathcal{F}$ and transfer matrix $\mathcal{V}$:

$$\mathcal{F} = \begin{bmatrix} (\alpha S + \beta R)\frac{I+D}{N} \\ 0 \end{bmatrix}, \quad \text{and } \mathcal{V} = \begin{bmatrix} (v+m)I \\ -vI + (m+w\eta)D \end{bmatrix}, \tag{C.1}$$

where $\mathcal{F} - \mathcal{V} = \left[\begin{smallmatrix} \dot{I} \\ \dot{D} \end{smallmatrix}\right]$ holds true. Then we calculate the corresponding Jacobian matrices

with respect to I and D at DFE $\mathcal{E}$:

$$F = \alpha \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \text{and } V = \begin{bmatrix} v+m & 0 \\ -v & m+w\eta \end{bmatrix}, \tag{C.2}$$

and the reproduction number $\mathscr{R}_c$ is the dominant eigenvalue of matrix FV^{-1}:

$$\begin{aligned} FV^{-1} =& \frac{\alpha}{(v+m)(w\eta+m)} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w\eta+m & 0 \\ v & v+m \end{bmatrix} \\ =& \frac{\alpha}{(v+m)(w\eta+m)} \begin{bmatrix} v+w\eta+m & v+m \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

and thus

$$\mathscr{R}_c(v,w) = \frac{\alpha(v+w\eta+m)}{(v+m)(w\eta+m)},$$

as a function of intervention v and w. □

Proof of Proposition 4.3.

Local asymptotic stability: Throughout this proof, we let state $x = (S, I, D, R)$ represent the vector of compartment sizes, $\dot{x} = f(x)$ represent the system of dynamics (4.1)-(4.4), $P = \frac{I+D}{S+I+D+R}$ represent the prevalence rate (i.e., the fraction of infected people among the total population), and $P_S = \frac{\partial P}{\partial S}$ (and similar for P_I, P_D, and P_R). Then,

$$\frac{\partial f}{\partial x} = \begin{bmatrix} -\mu - \alpha P - \alpha S P_S & -\alpha S P_I & -\alpha S P_D & -\alpha S P_R \\ \alpha P + (\alpha S + \beta R) P_S & (\alpha S + \beta R) P_I - m - v & (\alpha S + \beta R) P_D & (\alpha S + \beta R) P_R + \beta P \\ 0 & v & -m - w\eta & 0 \\ -\beta R P_S & -\beta R P_I & w\eta - \beta R P_D & -\beta P - \beta R P_R - \mu \end{bmatrix}, \tag{C.3}$$

$$\frac{\partial P}{\partial x} = \begin{bmatrix} P_S \\ P_I \\ P_D \\ P_R \end{bmatrix} = \frac{1}{(S+I+D+R)^2} \begin{bmatrix} -(I+D) \\ S+R \\ S+R \\ -(I+D) \end{bmatrix} \tag{C.4}$$

At DFE $x = \mathcal{E} = (S^* = \frac{\Lambda}{\mu}, 0, 0, 0)$, $P = 0$,

$$\left.\frac{\partial P}{\partial x}\right|_{x=\mathcal{E}} = \begin{bmatrix} 0 \\ 1/S^* \\ 1/S^* \\ 0 \end{bmatrix}, \text{ and } \left.\frac{\partial f}{\partial x}\right|_{x=\mathcal{E}} = \begin{bmatrix} -\mu & -\alpha & -\alpha & 0 \\ 0 & \alpha - m - v & \alpha & 0 \\ 0 & v & -m - w\eta & 0 \\ 0 & 0 & w\eta & -\mu \end{bmatrix}. \quad \text{(C.5)}$$

Matrix $\left.\frac{\partial f}{\partial x}\right|_{x=\mathcal{E}}$ has eigenvalues λ_i, $i = 1, \cdots, 4$: $\lambda_1 = \lambda_4 = -\mu$, λ_2 and λ_3 are roots to $|A - \lambda I| = 0$ where

$$A = \begin{bmatrix} \alpha - m - v & \alpha \\ v & -m - w\eta \end{bmatrix}, \text{ and then } \lambda_{2,3} = \frac{Tr(A) \pm \sqrt{(Tr(A))^2 - 4Det(A)}}{2}, \quad \text{(C.6)}$$

where

$$Det(A) = -(\alpha - m - v)(m + w\eta) - \alpha v = (m + v)(m + w\eta)(1 - \mathscr{R}_c), \quad \text{(C.7)}$$

$$Tr(A) = \alpha - v - w\eta - 2m. \quad \text{(C.8)}$$

By Theorem 4.7 in (117), $\mathcal{E}$ is locally asymptotically stable if and only if real part of all eigenvalues $\lambda_i, i = 1, \cdots, 4$ are negative. By (C.6) it is equivalent to have $Det(A) > 0$ and $Tr(A) < 0$, which reduces to $\mathscr{R}_c < 1$ because $Det(A) > 0$ holds true by (C.7), and $Tr(A) < 0$ holds true by the following:

$$\mathscr{R}_c = \frac{\alpha(v + w\eta + m)}{(v + m)(w\eta + m)} < 1 \implies \alpha < v + m \text{ and } \alpha < w\eta + m \quad \text{(C.9)}$$

$$\implies 2\alpha < v + w\eta + 2m \quad \text{(C.10)}$$

$$\implies Tr(A) = \alpha - v - w\eta - 2m < -\alpha < 0. \quad \text{(C.11)}$$

This concludes the proof for the local stability.

Global asymptotic stability:

Since $\alpha \geq \beta$ by Assumption 4.1, we have $\alpha S + \beta R \leq \alpha(S + R)$ and therefore

$\frac{\alpha S+\beta R}{N} \leq \alpha$. Then, the system of equations (4.2)-(4.3) leads to the following inequality:

$$\begin{bmatrix} \dot{I} \\ \dot{D} \end{bmatrix} \leq \begin{bmatrix} \alpha - m - v & \alpha \\ v & -m - w\eta \end{bmatrix} \begin{bmatrix} I \\ D \end{bmatrix} = (F - V) \begin{bmatrix} I \\ D \end{bmatrix}, \tag{C.12}$$

where matrices F and V are defined in (C.2).

Next we can easily check the following properties of matrices F and V:

- $F \geq 0$
- V has the form $V = sI - B$ with $B \geq 0$
- $V^{-1} = \frac{1}{(v+m)(w\eta+m)} \begin{bmatrix} w\eta + m & 0 \\ v & v + m \end{bmatrix} \geq 0$

By Lemmas 1 and 2 in (225), all eigenvalues of $(F - V)$ have negative real parts if and only if $\mathscr{R}_c < 1$. It implies that the linear ODE system

$$\begin{bmatrix} \dot{\tilde{I}} \\ \dot{\tilde{D}} \end{bmatrix} = (F - V) \begin{bmatrix} \tilde{I} \\ \tilde{D} \end{bmatrix} \tag{C.13}$$

is globally stable when $\mathscr{R}_c < 1$. Therefore, $(\tilde{I}(t), \tilde{D}(t)) \to (0, 0)$ as $t \to \infty$. By comparison principle (124), given the same initial condition (I_0, D_0), we have $I(t) \leq \tilde{I}(t), D(t) \leq \tilde{D}(t), \forall t \geq 0$, which implies that $(I(t), D(t)) \to (0, 0)$ as $t \to \infty$ for the nonlinear system (4.2)-(4.3). Substituting $I = D = 0$ to equations (4.1) and (4.4), we have that $S(t) \to S^*$ and $R(t) \to 0$ as $t \to \infty$. Therefore, $(S(t), I(t), D(t), R(t)) \to \mathcal{E} = (\frac{\lambda}{\mu}, 0, 0, 0)$ as $t \to \infty$. That is, the disease-free equilibrium $\mathcal{E}$ is globally asymptotically stable if $\mathscr{R}_c < 1$. □

Proof of Proposition 4.4: Let $P = \frac{I+D}{S+I+D+R}$ present the prevalence rate. By

solving equations $\frac{dS}{dt} = \frac{dI}{dt} = \frac{dD}{dt} = \frac{dR}{dt} = 0$, we have

$$S = \frac{\Lambda}{\mu + \alpha P} \tag{C.14}$$

$$\alpha S + \beta R = \frac{v+m}{P} \cdot I \tag{C.15}$$

$$I = \frac{m + w\eta}{v} \cdot D \tag{C.16}$$

$$R = \frac{w\eta}{\mu + \beta P} \cdot D. \tag{C.17}$$

Moreover, by plugging (C.16) and (C.17) to (C.15) we have

$$\alpha S + \beta R = \frac{v+m}{P} \cdot \frac{m + w\eta}{v} \cdot \frac{\mu + \beta P}{w\eta} R \tag{C.18}$$

$$\implies S = \frac{1}{\alpha}\left(\frac{v+m}{P} \cdot \frac{m + w\eta}{v} \cdot \frac{\mu + \beta P}{w\eta} - \beta\right) R \tag{C.19}$$

Also note that $P(S+I+D+R) = I+D$; by plugging in the results of (C.14), (C.16), (C.17), and (C.19), we have

$$0 = (P-1)(I+D) + PR + PS \tag{C.20}$$

$$= \left[\left(\frac{m + w\eta}{v} + 1\right) \cdot \frac{\mu + \beta P}{w\eta}(P-1) + P\right] R + P\left[\frac{1}{\alpha}\left(\frac{v+m}{P} \cdot \frac{m + w\eta}{v} \cdot \frac{\mu + \beta P}{w\eta} - \beta\right) R\right] \tag{C.21}$$

Multiply by $\frac{\alpha}{R}$ and then we get a quadratic equation of term P:

$$a_2 P^2 + a_1 P + a_0 = 0, \tag{C.22}$$

where

$$a_2 = \frac{\alpha\beta(v + w\eta + m)}{vw\eta} = \beta\frac{(v+m)(w\eta+m)}{vw\eta}\mathscr{R}_c > 0 \tag{C.23}$$

$$a_1 = \alpha\frac{v + w\eta + m}{vw\eta}(\mu - \beta) + \alpha + \frac{(v+m)(w\eta+m)}{vw\eta}\beta - \beta$$

$$= (\alpha - \beta) + \frac{(v+m)(w\eta+m)}{vw\eta}[\mathscr{R}_c\mu + \beta(1 - \mathscr{R}_c)] \tag{C.24}$$

$$a_0 = \frac{\mu}{vw\eta}(-(v + w\eta + m)\alpha + (v+m)(w\eta+m)) = \frac{\mu(v+m)(w\eta+m)}{vw\eta}(1 - \mathscr{R}_c), \tag{C.25}$$

and the root $P = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0a_2}}{2a_2}$. There exists endemic equilibrium if and only if there exists only one root $0 < P \leq 1$.

(1) When $\mathscr{R}_c > 1$. We know that $a_2 > 0$ by (C.23) and $a_0 < 0$ by (C.25). In this case, there is always one positive root no matter whether $a_1 < 0$ or $a_1 > 0$ because $\sqrt{a_1^2 - 4a_0a_2} > |a_1|$. Then it is sufficient to check

$$\begin{aligned} &a_1^2 - 4a_0a_2 \geq 0, \text{ and} \\ &-a_1 + \sqrt{a_1^2 - 4a_0a_2} < 2a_2 \iff a_0 + a_1 + a_2 > 0. \end{aligned}$$

The former is trivial, and the latter can be verified as follows:

$$\begin{aligned} a_0 + a_1 + a_2 =& \frac{(v+m)(w\eta+m)}{vw\eta}\left[\mu(1-\mathscr{R}_c) + \mu\mathscr{R}_c + \beta(1-\mathscr{R}_c)\right] \\ &+ (\alpha - \beta) + \beta\frac{(v+m)(w\eta+m)}{vw\eta}\mathscr{R}_c \\ =& \frac{(v+m)(w\eta+m)}{vw\eta}(\mu+\beta) + (\alpha-\beta) > 0. \end{aligned} \tag{C.26}$$

Therefore, there exists a unique EE when $\mathscr{R}_c > 1$.

(2) When $\mathscr{R}_c < 1$. We have $\alpha_0, \alpha_1, \alpha_2 > 0$. Then both roots $P = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0a_2}}{2a_2}$ are either negative real numbers or complex numbers with negative real parts; neither case implies a meaningful solution of prevalence P at an EE. In other words, there exists no EE in such a case. □

Proof of Corollary 4.1. Consider the equilibrium disease prevalence $P = \frac{I+D}{S+I+D+R}$ as a function of reproduction number $\mathscr{R}_c$. A necessary and sufficient condition of backward bifurcation (152) is:

$$\left.\frac{\partial P}{\partial \mathscr{R}_c}\right|_{\mathscr{R}_c=1, P=0} < 0.$$

Recall that P is the root of equation (C.22). Taking the derivative with respect to $\mathscr{R}_c$ at both sides, we have

$$\frac{da_2}{d\mathscr{R}_c}P^2 + a_2 \cdot 2P \cdot \frac{dP}{d\mathscr{R}_c} + \frac{da_1}{d\mathscr{R}_c}P + a_1\frac{dP}{d\mathscr{R}_c} - \frac{(v+m)(w\eta+m)}{w\eta} = 0,$$

at the bifurcation point ($\mathscr{R}_c = 1, P = 0$),

$$a_1 \frac{dP}{d\mathscr{R}_c} = \frac{(v+m)(w\eta+m)}{w\eta} \iff \frac{dP}{d\mathscr{R}_c} = \frac{(v+m)(w\eta+m)}{a_1 w\eta}.$$

To let $\frac{\partial P}{\partial \mathscr{R}_c}\Big|_{\mathscr{R}_c=1,P=0} < 0$, it is equivalent to have $a_1 < 0$, i.e.,

$$\beta > \alpha + \frac{(v+m)(w\eta+m)}{vw\eta}(\mu\mathscr{R}_c + \beta(1-\mathscr{R}_c)).$$

□

Proof of Proposition 4.5.

We can show that the optimal control exists for problem (4.10), and the existence can be generalized for any cost function which is convex in control variable u. First we modify the the formulation to the form that is suitable for applying Filippov-Cesari Theorem (Theorem 2.8 in (190)) as follows.

We define an auxiliary state variable $Y(t) = -(I(t) + D(t))$. With the augmented states $x = (S, I, D, R, Y)$, we have the following formulation:

Objective function:	$\max \int_0^{t_f} (-f_0(x,u,t))dt$
	where $f_0(x,u,t)$ is assumed to be convex in control variable u,
System dynamics:	$\dot{x} = f(x,u) := (f_1(x,u), f_2(x,u), f_3(x,u), f_4(x,u), f_5(x,u))$,
	where $f_i(x,u)$ for $i = 1,2,3,4$ are defined in (4.1)-(4.4),
	and $f_5(x,u) = -f_2(x,u) - f_3(x,u)$,
Initial conditions:	$x(0) = (S_0, I_0, D_0, R_0, -(I_0 + D_0))$
Terminal conditions:	$x_i(t_f)$ free for $i = 1,2,3,4$;
	$x_5(t_f) \geq -\rho^{target}$.

Now we are ready to apply the Filippov-Cesari Theorem, or Theorem 2.8 in (190). Define the set

$$N(x,U,t) := \{(-f_0(x,u,t) + \gamma, f_1(x,u,t), \cdots, f_5(x,u,t)) : \gamma \leq 0, u \in U\}.$$

It is sufficient to verify the following conditions:

(i) $N(x, U, t)$ is convex for each (x, t)

(ii) U is closed and bounded

(iii) There exists an admissible pair $(x(t), u(t))$

(iv) There exists a number b such that $\|x(t)\| \leq b$ for all $t \in [0, t_f]$ and all admissible pairs $(x(t), u(t))$.

(i) We first check the convexity of set $N(x, U, t)$. Since (x, t) is given, we suppress the dependency of x and t in the notation. We want to check the convexity of set $M(U) = \{(g_0(u)+\gamma, g_1(u), \cdots, g_5(u)) : \gamma \leq 0, u \in U\}$ where $g_0(u)$ is concave (because $g_0 = -f_0$), $g_i(u), i = 1, \cdots, 5$ are in u. Let y_1 and y_2 be two arbitrary points in set $M_{x,t}(U)$, i.e.,

$$y_1 = (g_0(u_1) + \gamma_1, g_1(u_1), g_2(u_1), g_3(u_1), g_4(u_1)), \text{ for some } \gamma_1 \leq 0, u_1 \in U,$$
$$y_2 = (g_0(u_2) + \gamma_2, g_1(u_2), g_2(u_2), g_3(u_2), g_4(u_2)), \text{ for some } \gamma_2 \leq 0, u_2 \in U,$$

Note that $U = [0, 1] \times [0, 1]$ and is convex. For any $\lambda \in [0, 1]$, let $u' = \lambda u_1 + (1-\lambda)u_2 \in U$, and

$$\begin{aligned}
&\lambda y_1 + (1 - \lambda)y_2 \\
=&(\lambda(g_0(u_1) + \gamma_1) + (1 - \lambda)(g_0(u_2) + \gamma_2), \\
&\lambda g_1(u_1) + (1 - \lambda g_1(u_2)), \cdots, \lambda g_5(u_1) + (1 - \lambda g_5(u_2))) \\
=&(\lambda(g_0(u_1) + \gamma_1) + (1 - \lambda)(g_0(u_2) + \gamma_2), g_1(u'), \cdots, g_5(u'))
\end{aligned}$$

where the second equation follows from the linearity of g_i for $i = 1, \cdots, 4$, and the last inequality follows from the concavity of function g_0 with respect to control u. Note that

$$\lambda g_0(u_1) + (1 - \lambda)g_0(u_2) + \lambda\gamma_1 + (1 - \lambda)\gamma_2 \leq g_0(u') + \lambda\gamma_1 + (1 - \lambda)\gamma_2$$

since g_0 is concave in u. Then the left-hand-side can be rewritten as

$$\lambda g_0(u_1)+(1-\lambda)g_0(u_2)+\lambda\gamma_1+(1-\lambda)\gamma_2 = g_0(u')+\gamma' \text{ for some } \gamma' \leq \lambda\gamma_1+(1-\lambda)\gamma_2 \leq 0.$$

Thus, $\lambda y_1+(1-\lambda)y_2 \in N(x,U,t)$ for any $\lambda \in [0,1]$, implying that $N(x,U,t)$ is convex for all choices of (x,t).

(ii) $U = [0,1] \times [0,1]$ and thus is closed and bounded.

(iii) Let $\hat{x}(t)$ be the solution to the system (4.1)-(4.6) with control $\hat{u}(t) \equiv 0$ for all t. $(\hat{x}, \hat{u})$ is an admissible pair.

(iv) As shown earlier in Proposition 4.1, the size the total population is always bounded. An thus, let $b = 2\max\{\Lambda/\mu, N_0\}$ and then we have $\|x(t)\| \leq |S(t)| + |I(t)| + |D(t)| + |R(t)| + |I(t)+D(t)| \leq 2N(t) \leq b$, which completes the proof. □

Proof of Proposition 4.6. Calculate the gradient of Hamiltonian $H(x^*, u, \lambda, t)$ with respect to control variables:

$$\begin{aligned}\frac{\partial H}{\partial v} &= \langle \lambda, g_1(x^*)\rangle = \lambda_0(c_s^r(S^*+I^*+R^*)+c_s^a I^*) + (\lambda_3-\lambda_2)I^* \\ \frac{\partial H}{\partial w} &= \langle \lambda, g_2(x^*)\rangle = \lambda_0 c_d D^* + (\lambda_4-\lambda_3)\eta D^*\end{aligned}$$

Notice that variable Z does not contribute to the equations of system dynamics, direct computation from (4.16) shows $\dot{\lambda}_0 = 0$; by the transversality condition, we have $\lambda_0(t) = 1$ for all t. Then we define the switching functions $\Phi_1 = \frac{\partial H}{\partial v}$ and $\Phi_2 = \frac{\partial H}{\partial w}$, which completes the proof. □

Proof of Proposition 4.7:

Case (1). Consider a simplified model with only S, I, and D compartments and the optimal control problem (4.12) with state $x = (Z, S, I, D)$ and

$$f(x) = \begin{bmatrix} 0 \\ \Lambda - \mu S - \alpha S\frac{I+D}{N} \\ \alpha S\frac{I+D}{N} - mI \\ -mD \end{bmatrix}, \; g_1(x) = \begin{bmatrix} c_s^a(S+I)+c_s^r I \\ 0 \\ -I \\ I \end{bmatrix}, \; g_2(x) = \begin{bmatrix} c_d D \\ 0 \\ 0 \\ -\eta D \end{bmatrix},$$

and $N = S + I + D$. Suppose there exists an singular arc on certain time interval $[t_1, t_2]$, and then switching functions vanish:

$$\Phi_1(t) = \langle \lambda, g_1 \rangle \equiv 0, \quad \Phi_2(t) = \langle \lambda, g_2 \rangle \equiv 0, \quad \forall t \in [t_1, t_2]. \tag{C.27}$$

Moreover,

$$\dot{\Phi}_1(t) = \dot{\Phi}_2(t) = 0, \quad \forall t \in [t_1, t_2].$$

To calculate these quantities, we use the mathematical tool called the *Lie bracket* of vector fields (135), which is defined as follows:

$$[f, g](x) = \frac{d}{dx} g(x) \cdot f(x) - \frac{d}{dx} f(x) \cdot g(x).$$

The properties of the lie bracket in the following lemma can be easily verified by direct calculations (also see Proposition 2 in (129)).

Lemma C.1 *The lie bracket* $[\cdot, \cdot]$ *has the following properties:*

- $[f, f] = 0$ *and* $[f, g] = -[g, f]$,
- *Let* λ *be the co-state vector corresponding to the problem* (4.12)*. Define function* $\Phi(t) = \langle \lambda, h \rangle$*. Then*

$$\dot{\Phi}(t) = \langle \lambda, [f + g_1 v + g_2 w, h] \rangle.$$

By Lemma C.1, we have

$$\dot{\Phi}_1(t) = \langle \lambda, [f + g_1 v + g_2 w, g_1] \rangle = \langle \lambda, [f, g_1] \rangle + \langle \lambda, [g_2, g_1] \rangle w = 0,$$
$$\dot{\Phi}_2(t) = \langle \lambda, [f + g_1 v + g_2 w, g_2] \rangle = \langle \lambda, [f, g_2] \rangle + \langle \lambda, [g_1, g_2] \rangle v = 0,$$

where

$$[f, g_1] = \begin{bmatrix} c_s^a(\Lambda - mI - \mu S) + c_s^r(-mI + \alpha S \frac{I+D}{N}) \\ 0 \\ -\alpha S \frac{I+D}{N} \\ \alpha S \frac{I+D}{N} \end{bmatrix},$$

$$[f, g_2] = \begin{bmatrix} -c_d Dm \\ -\alpha\eta D\frac{S^2}{N^2} \\ \alpha\eta D\frac{S^2}{N^2} \\ 0 \end{bmatrix}, \text{ and } [g_1, g_2] = \begin{bmatrix} c_d I \\ 0 \\ 0 \\ -\eta I \end{bmatrix}.$$

As $g_2 \propto [g_1, g_2]$ and $\Phi_2 = \langle \lambda, g_2 \rangle \equiv 0$, we have $\langle \lambda, [g_1, g_2] \rangle = 0$. Then $\dot{\Phi}_1(t)$ and $\dot{\Phi}_2(t)$ can be simplified to

$$\dot{\Phi}_1(t) = \langle \lambda, [f, g_1] \rangle \equiv 0, \text{ and } \dot{\Phi}_2(t) = \langle \lambda, [f, g_2] \rangle \equiv 0. \tag{C.28}$$

By (C.27) and (C.28) , λ has zero inner product with g_1, g_2, $[f, g_1]$, and $[f, g_2]$. However, these four vectors are linear independent, which implies $\lambda \equiv 0$. But $\lambda_0 \equiv 1$ as shown earlier. Therefore, control v and w cannot be simultaneously singular.

Case (2). Consider the full model (4.12) when $\alpha = \beta$. Then compartments S and R are indeed indifferent and can be combined as one compartment. As a result, the full model is equivalent to the optimal control problem (C.27) as in Case (1) except for a different $g_2'(x) = (c_d D, \eta D, 0, -\eta D)^\top$.

Following the same calculation as in the proof of Case (1), we have

$$[f, g_2'] = \begin{bmatrix} -c_d mD \\ -(m-\mu)\eta D + \alpha\eta D\frac{D+I-S}{N} \\ -\alpha\eta D\frac{D+I-S}{N} \\ 0 \end{bmatrix}, \ [g_1, g_2'] = \begin{bmatrix} c_d I - c_s \eta D \\ \eta I \\ 0 \\ -\eta I \end{bmatrix}$$

Note that

$$[g_1, g_2'] = \frac{I}{D} g_2' - c_s^a \eta D (1, 0, 0, 0)^\top = \frac{I}{D} g_2' - c_s^a \eta D (k_1 g_1 + k_2 [f, g_1]) \tag{C.29}$$

for some k_1 and k_2 that solve:

$$\begin{cases} k_1 I + k_2 \alpha S \frac{I+D}{N} = 0 \\ k_1 [c_s^a (S + I) + c_s^r I] + k_2 [c_s^a (\Lambda - mI - \mu S) + c_s^r (-mI + \alpha S \frac{I+D}{N})] = 1 \end{cases} \tag{C.30}$$

Thus,

$$\begin{aligned}0 \equiv &\dot{\Phi}_1 = \langle \lambda, [f, g_1] \rangle + \langle \lambda, [g_2', g_1] \rangle w \\ =&\langle \lambda, [f, g_1] \rangle - \langle \lambda, g_2' \rangle \frac{I}{D} + \langle \lambda, g_1 \rangle c_s \eta D k_1 + \langle \lambda, [f, g_1] \rangle c_s \eta D k_2 \\ =&\langle \lambda, [f, g_1] \rangle (1 + c_s \eta D k_2) \end{aligned} \tag{C.31}$$

Clearly $1+c_s\eta D k_2 \neq 0$ for k_2 value that solves (C.30). Then (C.31) implies $\langle \lambda, [f, g_1] \rangle \equiv 0$. Moreover, $\langle \lambda, [g_1, g_2'] \rangle \equiv 0$ by (C.29), and thus $\dot{\Phi}_2(t) = \langle \lambda, [f, g_2'] \rangle \equiv 0$. Therefore, λ has zero inner product with g_1, g_2', $[f, g_1]$, and $[f, g_2']$, which are indeed linear independent. This implies $\lambda \equiv 0$, which contradicts with $\lambda_0 \equiv 1$ and completes the proof.

□

Proof of Proposition 4.8. For problem (4.19), Hamiltonian is defined as $H = \langle \lambda, f + gw \rangle$ where $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ are the costates satisfying

$$\dot{\lambda} = -\frac{\partial H}{\partial x} = -\left\langle \lambda, \frac{\partial f(x^*)}{\partial x} \right\rangle + \left\langle \lambda, \frac{\partial g(x^*)}{\partial x} \right\rangle w^*, \tag{C.32}$$

with transversality conditions

$$\lambda_0(t_f) = 1,\ \lambda_1(t_f) = \lambda_4(t_f) = 0,\ \lambda_2(t_f) = \lambda_3(t_f) = k \tag{C.33}$$

for some $k \geq 0$ and $k = 0$ if $I^*(t_f) + D^*(t_f) < \rho^{target}$.

We observe that $\lambda_0(t) \equiv 1$ since $\dot{\lambda_0} = 0$ and $\lambda_0(t_f) = 1$. Then the switching function is:

$$\Phi = \frac{\partial H(x^*)}{\partial w} = \langle \lambda, g \rangle = D^*(c_d + \eta(\lambda_4 - \lambda_3)).$$

Suppose there exists a singular arc. Switching function $\Phi(t)$ vanishes on the singular arc, i.e., $\Phi(t) \equiv 0$. Since disease free equilibrium is not achieved, given $v^\circ \not\equiv 0$, compartment $I(t)$ and $D(t)$ cannot remain 0 over a time interval. Thus, $\Phi(t) \equiv 0$ implies

$$\lambda_4 - \lambda_3 = -\frac{c_d}{\eta}. \tag{C.34}$$

Furthermore, we have

$$\dot{\Phi} = \langle \lambda, [f, g] \rangle \equiv 0 \tag{C.35}$$

$$\ddot{\Phi} = \langle \lambda, [f, [f, g]] \rangle + \langle \lambda, [g, [f, g]] \rangle w \tag{C.36}$$

Direct calculation can show that

$$[f, g] = \begin{bmatrix} c_d(-mD + v^\circ I) + c_s^a \eta v^\circ D \\ -\frac{\alpha \eta D S}{N} \\ \frac{\alpha \eta D S}{N} - \frac{\beta \eta D(D+I-R)}{N} \\ -\eta v^\circ I \\ \frac{\beta \eta D(D+I-R)}{N} - ((m-\mu)D - v^\circ I) \end{bmatrix},$$

$$[g, [f, g]] = \begin{bmatrix} \eta(c_d m D + c_d v^\circ I - c_s^a \eta v^\circ D) \\ \frac{\alpha \eta^2 S D}{N} \\ \frac{\beta \eta^2 D(3D+I-R)}{N} - \frac{\alpha \eta^2 S D}{N} \\ -v^\circ \eta^2 I \\ -\frac{\beta \eta^2 D(3D+I-R)}{N} + (m-\mu)D + v^\circ I \end{bmatrix} = -\eta \cdot [f, g] + \begin{bmatrix} 2c_d \eta v^\circ I \\ 0 \\ \frac{2\beta \eta^2 D^2}{N} \\ -2\eta^2 v^\circ I \\ -\frac{2\beta \eta^2 D^2}{N} + 2v^\circ \eta^2 I \end{bmatrix}$$

Thus, on the singular arc,

$$\begin{aligned} \langle \lambda, [g, [f, g]] \rangle =& -\eta \langle \lambda, [f, g] \rangle + 2c_d \eta v^\circ I + \frac{2\beta \eta^2 D^2}{N}(\lambda_2 - \lambda_4) + 2v^\circ \eta^2 I(\lambda_4 - \lambda_3) \\ =& 2c_d \eta v^\circ I + \frac{2\beta \eta^2 D^2}{N}(\lambda_2 - \lambda_4) - 2v^\circ \eta c_d I \\ =& \frac{2\beta \eta^2 D^2}{N}(\lambda_2 - \lambda_3 + \lambda_3 - \lambda_4) \\ =& \frac{2\beta \eta^2 D^2}{N}(\lambda_2 - \lambda_3) + \frac{2\beta \eta c_d D^2}{N}, \end{aligned}$$

where the second and the fourth equality holds true by (C.34). Next, we determine the sign of $\lambda_2 - \lambda_3$ on the singular arc.

$$\begin{aligned} \dot{\lambda_2} - \dot{\lambda_3} =& -\left(\frac{\partial H}{\partial I} - \frac{\partial H}{\partial D} \right) \\ =& -(c_s^a + c_s^r)v^\circ + c_d w + (m + v^\circ)(\lambda_2 - \lambda_3) + \eta w(\lambda_4 - \lambda_3) \\ =& (m + v^\circ)(\lambda_2 - \lambda_3) - (c_s^a + c_s^r)v^\circ + w\Phi \end{aligned}$$

with terminal condition $\lambda_2(t_f) - \lambda_3(t_f) = 0$.

Consider a reverse process $\lambda_{23}(\tau) = \lambda_2(t_f - \tau) - \lambda_3(t_f - \tau)$ with screening rate $\bar{v}^\circ(\tau) = v^\circ(t_f - \tau)$, for $\tau \in [0, t_f]$. Then we have

$$\dot{\lambda_{23}}(\tau) = -(m+v)\lambda_{23}(\tau) + (c_s^a + c_s^r)\bar{v}^\circ(\tau) - w(t_f - \tau)\Phi(t_f - \tau)$$

with initial condition $\lambda_{23}(0) = 0$. Note that $w\Phi \leq 0$ always holds true because $w = 0$ when $\Phi > 0$ and $w = 1$ when $\Phi < 0$. Therefore $\dot{\lambda_{23}} \geq -(m + \bar{v}^\circ)\lambda_{23}$. Let $\dot{y} = -(m + \bar{v}^\circ)y$ with $y(0) = 0$. By comparison theorem (Theorem 1.5.2 in (124)), we have $\lambda_{23}(t) \geq y(t) = 0$ for all $t \in [0, t_f]$, which implies

$$\lambda_2(t) - \lambda_3(t) \geq 0, \ t \in [0, t_f].$$

Therefore, $\frac{d}{du}\left[\frac{d^2}{dt^2}\left(\frac{\partial H}{\partial u}\right)\right] = \langle \lambda, [g,[f,g]] \rangle \geq \frac{2\beta\eta c_d D^2}{N} > 0$, which contradicts to the generalized Legendre-Clebsch condition at order $k = 1$, i.e., $-\frac{d}{du}\left[\frac{d^2}{dt^2}\left(\frac{\partial H}{\partial u}\right)\right] \geq 0$, and excludes the existence of singular control arc. □

Proof of Proposition 4.9. The optimization problem is equivalent to

$$\max_{w(t)\in[0,1]} \{-J(v^c) : (4.1) - (4.6)\}.$$

Define the Hamiltonian $H = \langle \lambda, f(x) + g_1(x)v^c + g_2(x)w \rangle$ with the costate $\lambda(t) = (\lambda_1(t), \cdots, \lambda_4(t))$, where

$$f(x) = \begin{bmatrix} \Lambda - \mu S - \alpha S\frac{I+D}{N} \\ (\alpha S + \beta R)\frac{I+D}{N} - mI \\ -mD \\ -\mu R - \beta R\frac{I+D}{N} \end{bmatrix}, \ g_1(x) = \begin{bmatrix} 0 \\ -I \\ I \\ 0 \end{bmatrix}, \ g_2(x) = \begin{bmatrix} 0 \\ 0 \\ -\eta D \\ \eta D \end{bmatrix}. \tag{C.37}$$

We want to examine the effect of the constant screening rate v^c on the objective value $-J(v^c)$. The subderivative of objective value $-J(\cdot)$ at v^c, denoted by h, is as follows (see pg. 217 of (190)),:

$$h = \int_0^{t_f} \frac{\partial H(x^*, w^*, \lambda, t)}{\partial v^c} dt = \int_0^{t_f} I^*(\lambda_3 - \lambda_2) dt.$$

Note that costates $\lambda(t)$ satisfy $\dot{\lambda} = -\frac{\partial H}{\partial x}$ with transversality conditions $\lambda_1(t_f) = \lambda_4(t_f) = 0, \lambda_2(t_f) = \lambda_3(t_f) = 1$. By direct calculation,

$$\dot{\lambda_3} - \dot{\lambda_2} = (m + v^c)(\lambda_3 - \lambda_2) - w\eta(\lambda_4 - \lambda_3).$$

Note $w\eta(\lambda_4 - \lambda_3) \geq 0$ because $w = 1$ when switching function $\eta(\lambda_4 - \lambda_3) > 0$ and $w = 0$ when $\eta(\lambda_4 - \lambda_3) < 0$. Thus $\dot{\lambda_3} - \dot{\lambda_2} \leq (m + v)(\lambda_3 - \lambda_2)$. Since $\lambda_3(t_f) - \lambda_2(t_f) = 0$, by time reversion and the comparison theorem, we have $\lambda_3(t) - \lambda_2(t) \geq 0$ for all $t \in [0, t_f]$. Thus, subgradient

$$h = \int_0^{t_f} I(\lambda_3 - \lambda_2)dt \geq 0,$$

which implies

$$-J(v) \geq -J(v^c) + h(v - v^c) \geq -J(v^c), \ \forall v \geq v^c,$$

and therefore $J(v) \leq J(v^c), \forall v \geq v^c$. □

Proof of Proposition 4.10. Since the total population $N(t)$ remains a constant N_0 by assumption, the model formulation can be scaled by a factor of $\frac{1}{N_0}$. Then state $x(t) = (S(t), I(t), D(t), R(t)$ represents a population distribution which sums to one. We also let $\overline{w}$ represent the actual number of treatment, i.e., $\overline{w}(t) = w(t)D(t)$, $\overline{w}(t) \leq D(t)$. Then the system dynamics are as follows:

$$\dot{S} = \mu - \mu S - \alpha S(I + D) \tag{C.38}$$

$$\dot{I} = (\alpha S + \beta R)(I + D) - \mu I - vI \tag{C.39}$$

$$\dot{D} = vI - \mu D - \eta\overline{w} \tag{C.40}$$

$$\dot{R} = \eta\overline{w} - \mu R - \beta R(I + D) \tag{C.41}$$

$$v \in [0, 1], \ \overline{w} \leq D$$

For the uninfected population (i.e., the susceptibles and the recovered), we define

the state $y = (S, R)$ with the system dynamics for any given treatment number $\overline{w}$:

$$h_{\overline{w}}(y) = \begin{bmatrix} \dot{S} \\ \dot{R} \end{bmatrix} = \begin{bmatrix} \mu - \mu S - \alpha S(I + D) \\ \eta\overline{w} - \mu R - \beta R(I + D) \end{bmatrix} = \begin{bmatrix} \mu - \mu S - \alpha S(1 - S) + \alpha SR \\ \eta\overline{w} - \mu R - \beta R(1 - R) + \beta RS \end{bmatrix} \tag{C.42}$$

where the second inequality follows from the fact that $S + I + D + R = 1$.

Next, we introduce the definition of *quasimonotone nondecreasing property* (124) (also referred to as "cooperative" in the literature) that plays a key role in establishing differential inequality results.

Definition C.1 (Quasimonotonicity, Laskshmikantham et al. (2015)) *Function $g(t, x) : \mathbb{R}_+ \times \mathbb{R}^n \mapsto \mathbb{R}^n$ is said to be quasimonotone nondecreasing if for any given t,*

$$x \leq y, x_i = y_i \implies f_i(t, x) \leq f_i(t, y) \quad (i = 1, \cdots, n).$$

It is straightforward to check that $h_{\overline{w}}(y)$ is indeed a quasimonotone nondecreasing function. Also note that $h_{\overline{w}_1}(y) \leq h_{\overline{w}_2}(y)$ because $\overline{w}_1 \leq \overline{w}_2$.

We are now ready to apply Theorem 1.5.4 of (124) to show the inequalities $S_1(t) \leq S_2(t)$ and $R_1(t) \leq R_2(t)$ as follows. Since state $y(t) = (S(t), R(t))$ is differentiable, Dini derivative of $y(t)$ at time t is the usual derivative $\dot{y} = h$. Given the same initial condition $y_1(0) = y_2(0)$, $\dot{y}_i = h_{\overline{w}_i}(y_i)$ for $i = 1, 2$, and quasimonotone nondecreasing function $h_{\overline{w}_2}$, by Theorem 1.5.4 of (124), $h_{\overline{w}_1(t)}(y) \leq h_{\overline{w}_2(t)}(y)$ implies that $y_1(t) \leq y_2(t)$ component-wisely for all $t \geq 0$. That is, $S_1(t) \leq S_2(t)$, $R_1(t) \leq R_2(t)$, and thus $I_1(t) + D_1(t) \geq I_2(t) + D_2(t)$ for all time t. □

Proof of Proposition 4.11. With the assumption of stationary population, we consider the system dynamics (C.38)-(C.41) with the state $x = (S, I, D, R)$ now representing the population distribution. Budget B is scaled accordingly and represents the annual budget per person.

Case (1). When $B \leq c_d D$, i.e., budget is not sufficient to treat all diagnosed patients. By Proposition 4.10, more people receiving treatment (i.e., a higher value of $\overline{w}(t) =$

$w(t)D(t))$ always leads to lower $I(t) + D(t)$, regardless of the value of screening rate $v(t)$. Therefore, in this case, it is optimal to increase treatment number to the maximum possible value, i.e., by allocating all the budget to treatment and none for screening.

Case (2). When $B > c_d D$. By the result of Case (1), budgets are first allocated for treatment until all diagnosed patients are covered. To show that the remaining budget should be spent on screening, it is sufficient to show that at the treatment rate $w(t) = 1$, increasing the number of screenings will reduce the prevalence $I(t) + D(t)$.

Consider the susceptible, diagnosed, and recovered population. We define state $y = (S, D, R)$ with the following dynamics

$$l_{\overline{v}}(y) = \begin{bmatrix} \dot{S} \\ \dot{D} \\ \dot{R} \end{bmatrix} = \begin{bmatrix} \Lambda - \mu S - \alpha S(1-S) + \alpha SR \\ \overline{v} - (m+\eta)D \\ \eta D - \mu R - \beta R(1-R) + \beta RS \end{bmatrix} \tag{C.43}$$

where $\overline{v}$ represents the actual number of diagnosed patients through the screening.

It is straightforward to check that $l_{\overline{v}}(y)$ is quasimonotone nondecreasing. Similar to the proof of Proposition 4.10, by applying Theorem 1.5.4 of (124), when $\overline{v}_1(t) \leq \overline{v}_2(t)$, we have $S_1(t) \leq S_2(t)$ and $R_1(t) \leq R_2(t)$, which implies $I_1(t) + D_1(t) \geq I_2(t) + D_2(t)$ for all t, leading to a lower objective value of (4.25).

Therefore, given that the treatment is saturated with the maximum rate, it is optimal to increase the number of screening as much as possible, i.e., by allocating all remaining budget to screening. □

Bibliography

[1] "Engauge digitizer." `http://digitizer.sourceforge.net/`, 2015.

[2] "UpToDate: Drug Information." `http://www.uptodate.com/home`, 2015.

[3] ABIM, "Medical professionalism in the new millennium: A physician charter," *Annals of Internal Medicine*, vol. 136, no. 3, p. 243, 2002.

[4] Adena, M., Houltram, J., Mulligan, S. P., Todd, C., and Malanos, G., "Modelling the cost effectiveness of rituximab in chronic lymphocytic leukaemia in first-line therapy and following relapse," *Pharmacoeconomics*, vol. 32, no. 2, pp. 193–207, 2014.

[5] Akan, M., Alagoz, O., Ata, B., Erenay, F. S., and Said, A., "A broader view of designing the liver allocation system," *Operations research*, vol. 60, no. 4, pp. 757–770, 2012.

[6] Alagoz, O., Maillart, L., Schaefer, A., and Roberts, M., "Determining the acceptance of cadaveric livers using an implicit model of the waiting list," *Operations Research*, vol. 55, no. 1, pp. 24–36, 2007.

[7] Alistar, S. S., Brandeau, M. L., and Beck, E. J., "REACH: A practical HIV resource allocation tool for decision makers," in *Operations Research and Health Care Policy*, pp. 201–223, Springer, 2013.

[8] Alistar, S. S., Long, E. F., Brandeau, M. L., and Beck, E. J., "HIV epidemic control: A model for optimal allocation of prevention and treatment resources," *Health care management science*, vol. 17, no. 2, pp. 162–181, 2014.

[9] Allen, L. J., Brauer, F., Van den Driessche, P., and Wu, J., *Mathematical epidemiology*. Springer, 2008.

[10] Altekruse, S., McGlynn, K., and Reichman, M., "Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005," *Journal of Clinical Oncology*, vol. 27, no. 9, pp. 1485–1491, 2009.

[11] Altman, E., *Constrained Markov decision processes*, vol. 7. CRC Press, 1999.

[12] Anderson, G. F., *Chronic care: Making the case for ongoing care*. Robert Wood Johnson Foundation, 2010.

[13] Andersson, K., Salomon, J., Goldie, S., and Chung, R., "Cost effectiveness of alternative surveillance strategies for hepatocellular carcinoma in patients with cirrhosis," *Clinical Gastroenterology and Hepatology*, vol. 6, no. 12, pp. 1418–1424, 2008.

[14] Arguedas, M., Chen, V., Eloubeidi, M., and Fallon, M., "Screening for hepatocellular carcinoma in patients with hepatitis C cirrhosis: A cost-utility analysis," *The American journal of gastroenterology*, vol. 98, no. 3, pp. 679–690, 2003.

[15] Arias, E., "United States life tables, 2007.," *Natl Vital Stat Rep*, vol. 59, no. 9, p. 1, 2011.

[16] Arias, E., "United States life tables, 2010," *Natl Vital Stat Rep*, vol. 63, no. 7, pp. 1–63, 2014.

[17] Atkins, D., Ross, D., and Kelley, M., "Acting in the face of uncertainty," *Annals of internal medicine*, vol. 161, no. 4, pp. 300–301, 2014.

[18] Ayer, T., Alagoz, O., and Stout, N., "OR Forum: A POMDP approach to personalize mammography screening decisions," *Operations Research*, vol. 60, no. 5, pp. 1019–1034, 2012.

[19] Ayer, T., Alagoz, O., Stout, N. K., and Burnside, E. S., "Heterogeneity in womens adherence and its role in optimal breast cancer screening policies," *Management Science*, vol. 62, no. 5, pp. 1339–1362, 2015.

[20] Bach, P. B., "Limits on medicare's ability to control rising spending on cancer drugs," *New England Journal of Medicine*, vol. 360, no. 6, pp. 626–633, 2009.

[21] Bach, P. B., "New math on drug cost-effectiveness," *New England Journal of Medicine*, vol. 373, no. 19, pp. 1797–1799, 2015.

[22] Bakare, E. A., Nwagwo, A., and Danso-Addo, E., "Optimal control analysis of an sir epidemic model with constant recruitment," *International Journal of Applied Mathematical Research*, vol. 3, no. 3, pp. 273–285, 2014.

[23] Bedossa, P. and Poynard, T., "An algorithm for the grading of activity in chronic hepatitis C," *Hepatology*, vol. 24, no. 2, pp. 289–293, 1996.

[24] Benson 3rd, A., Abrams, T., Ben-Josef, E., Bloomston, P., Botha, J., Clary, B., Covey, A., Curley, S., D'Angelica, M., Davila, R., and others, "Nccn clinical practice guidelines in oncology: Hepatobiliary cancers.," *Journal of the National Comprehensive Cancer Network: JNCCN*, vol. 7, no. 4, p. 350, 2009.

[25] Beusterien, K. M., Davies, J., Leach, M., Meiklejohn, D., Grinspan, J. L., O'Toole, A., and Bramham-Jones, S., "Population preference values for treatment outcomes in chronic lymphocytic leukaemia: a cross-sectional utility study," *Health and quality of life outcomes*, vol. 8, no. 1, p. 50, 2010.

[26] Bochud, P., Cai, T., Overbeck, K., Bochud, M., Dufour, J., Müllhaupt, B., Borovicka, J., Heim, M., Moradpour, D., Cerny, A., and others, "Genotype 3 is associated with accelerated fibrosis progression in chronic hepatitis C," *Journal of hepatology*, vol. 51, no. 4, pp. 655–666, 2009.

[27] Bodenheimer, T., Chen, E., and Bennett, H. D., "Confronting the growing burden of chronic disease: can the us health care workforce do the job?," *Health Affairs*, vol. 28, no. 1, pp. 64–74, 2009.

[28] Brandeau, M. L., Zaric, G. S., and Richter, A., "Resource allocation for control of infectious diseases in multiple independent populations: beyond cost-effectiveness analysis," *Journal of health economics*, vol. 22, no. 4, pp. 575–598, 2003.

[29] Brauer, F., Castillo-Chavez, C., and Castillo-Chavez, C., *Mathematical models in population biology and epidemiology*, vol. 40. Springer, 2001.

[30] Briat, C. and Verriest, E. I., "A new delay-SIR model for pulse vaccination," *Biomedical signal processing and control*, vol. 4, no. 4, pp. 272–277, 2009.

[31] Briggs, A., Sculpher, M., and Claxton, K., *Decision modelling for health economic evaluation*. Oxford university press, 2006.

[32] Brown, J. R., Byrd, J. C., Coutre, S. E., Benson, D. M., Flinn, I. W., Wagner-Johnston, N. D., Spurgeon, S. E., Kahl, B. S., Bello, C., Webb, H. K., and others, "Idelalisib, an inhibitor of phosphatidylinositol 3-kinase p110δ, for relapsed/refractory chronic lymphocytic leukemia," *Blood*, vol. 123, no. 22, pp. 3390–3397, 2014.

[33] Bruix, J. and Sherman, M., "Management of hepatocellular carcinoma: An update," *Hepatology*, vol. 53, no. 3, pp. 1020–1022, 2011.

[34] Bryson, A. E., *Applied optimal control: Optimization, estimation and control*. CRC Press, 1975.

[35] Buckley, G. J., Strom, B. L., and others, *Eliminating the Public Health Problem of Hepatitis B and C in the United States: Phase One Report*. National Academies Press, 2016.

[36] Burger, J. A., Tedeschi, A., Barr, P. M., Robak, T., Owen, C., Ghia, P., Bairey, O., Hillmen, P., Bartlett, N. L., Li, J., and others, "Ibrutinib as initial therapy for patients with chronic lymphocytic leukemia," *New England Journal of Medicine*, vol. 373, no. 25, pp. 2425–2437, 2015.

[37] Byrd, J. C., Brown, J. R., O'Brien, S., Barrientos, J. C., Kay, N. E., Reddy, N. M., Coutre, S., Tam, C. S., Mulligan, S. P., Jaeger, U., and others, "Ibrutinib versus ofatumumab in previously treated chronic lymphoid leukemia," *New England Journal of Medicine*, vol. 371, no. 3, pp. 213–223, 2014.

[38] Byrd, J. C., Furman, R. R., Coutre, S. E., Burger, J. A., Blum, K. A., Coleman, M., Wierda, W. G., Jones, J. A., Zhao, W., Heerema, N. A., and others, "Three-year follow-up of treatment-naive and previously treated patients with cll and sll receiving single-agent ibrutinib," *Blood*, pp. blood–2014, 2015.

[39] Cabrera, R. and Nelson, D., "Review article: The management of hepatocellular carcinoma," *Alimentary pharmacology & therapeutics*, vol. 31, no. 4, pp. 461–476, 2010.

[40] Castera, L., "Transient elastography and other noninvasive tests to assess hepatic fibrosis in patients with viral hepatitis," *Journal of viral hepatitis*, vol. 16, no. 5, pp. 300–314, 2009.

[41] Catovsky, D., Richards, S., Matutes, E., Oscier, D., Dyer, M., Bezares, R., Pettitt, A., Hamblin, T., Milligan, D., and Child, J., "Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): A randomised controlled trial," *The Lancet*, vol. 370, no. 9583, pp. 230–239, 2007.

[42] Centers for Disease Control and Prevention, "Chronic disease prevention and health promotion." `http://www.cdc.gov/chronicdisease/resources/publications/four-domains.htm`, 2016.

[43] Centers for Medicare and Medicaid Services, "Medicare physician fee schedule." `http://www.cms.gov/apps/physician-fee-schedule/`, 2015.

[44] Chen, Q., Ayer, T., Nastoupil, L. J., Rose, A. C., and Flowers, C. R., "Comparing the cost-effectiveness of rituximab maintenance and radioimmunotherapy consolidation versus observation following first-line therapy in patients with follicular lymphoma," *Value in Health*, vol. 18, no. 2, pp. 189–197, 2015.

[45] Cheung, Y., Thumboo, J., Gao, F., Ng, G., Pang, G., Koo, W., Sethi, V., Wee, J., and Goh, C., "Mapping the english and chinese versions of the functional assessment of cancer therapy–general to the EQ-5D utility index," *Value in health*, vol. 12, no. 2, pp. 371–376, 2009.

[46] Chhatwal, J., Alagoz, O., and Burnside, E., "Optimal breast biopsy decision-making based on mammographic features and demographic factors," *Operations research*, vol. 58, no. 6, pp. 1577–1591, 2010.

[47] Chhatwal, J., Ferrante, S., Brass, C., El Khoury, A., Burroughs, M., Bacon, B., Esteban-Mur, R., and Elbasha, E., "Cost-effectiveness of boceprevir in patients previously treated for chronic hepatitis C genotype 1 infection in the United States," *Value in Health*, vol. 16, no. 6, pp. 973–986, 2013.

[48] Chhatwal, J., Mathisen, M., and Kantarjian, H., "Are high drug prices for hematologic malignancies justified? a critical analysis," *Cancer*, vol. 121, no. 19, pp. 3372–3379, 2015.

[49] Chong, C., Gulamhussein, A., Heathcote, E., Lilly, L., Sherman, M., Naglie, G., and Krahn, M., "Health-state utilities and quality of life in hepatitis C patients," *The American journal of gastroenterology*, vol. 98, no. 3, pp. 630–638, 2003.

[50] Colli, A., Fraquelli, M., Casazza, G., Massironi, S., Colucci, A., Conte, D., and Duca, P., "Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: A systematic review," *The American journal of gastroenterology*, vol. 101, no. 3, pp. 513–523, 2006.

[51] Coon, J., Rogers, G., Hewson, P., Wright, D., Anderson, R., Jackson, S., Ryder, S., Cramp, M., and Stein, K., "Surveillance of cirrhosis for hepatocellular carcinoma: A cost–utility analysis," *British journal of cancer*, vol. 98, no. 7, pp. 1166–1175, 2008.

[52] Cousien, A., Tran, V. C., Deuffic-Burban, S., Jauffret-Roustide, M., Dhersin, J.-S., and Yazdanpanah, Y., "Dynamic modelling of hepatitis c virus transmission among people who inject drugs: a methodological review," *Journal of viral hepatitis*, vol. 22, no. 3, pp. 213–229, 2015.

[53] Curry, M., Forns, X., Chung, R., Terrault, N., Brown, R., Fenkel, J., Gordon, F., OLeary, J., Kuo, A., Schiano, T., and others, "Sofosbuvir and ribavirin prevent recurrence of HCV infection after liver transplantation: An open-label study," *Gastroenterology*, vol. 148, no. 1, pp. 100–107, 2015.

[54] D'Amico, G., Garcia-Tsao, G., and Pagliaro, L., "Natural history and prognostic indicators of survival in cirrhosis: A systematic review of 118 studies," *Journal of hepatology*, vol. 44, no. 1, pp. 217–231, 2006.

[55] Davis, G., Alter, M., El-Serag, H., Poynard, T., and Jennings, L., "Aging of hepatitis C virus (HCV)-infected persons in the United States: A multiple cohort model of HCV prevalence and disease progression," *Gastroenterology*, vol. 138, no. 2, pp. 513–521, 2010.

[56] Deo, S., Rajaram, K., Rath, S., Karmarkar, U. S., and Goetz, M. B., "Planning for hiv screening, testing, and care at the veterans health administration," *Operations Research*, vol. 63, no. 2, pp. 287–304, 2015.

[57] DeVol, R., *An unhealthy America: The economic burden of chronic disease–charting a new course to save lives and increase productivity and economic growth.* Milken Institute, 2007.

[58] Diekmann, O., Heesterbeek, J., and Roberts, M., "The construction of next-generation matrices for compartmental epidemic models," *Journal of the Royal Society Interface*, p. rsif20090386, 2009.

[59] Diekmann, O., Heesterbeek, H., and Britton, T., *Mathematical tools for understanding infectious disease dynamics.* Princeton University Press, 2012.

[60] Dienstag, J., Ghany, M., Morgan, T., Di Bisceglie, A., Bonkovsky, H., Kim, H., Seeff, L., Szabo, G., Wright, E., Sterling, R., and others, "A prospective study of the rate of progression in compensated, histologically advanced chronic hepatitis C," *Hepatology*, vol. 54, no. 2, pp. 396–405, 2011.

[61] Ditah, I., Ditah, F., Devaki, P., Ewelukwa, O., Ditah, C., Njei, B., Luma, H., and Charlton, M., "The changing epidemiology of hepatitis C virus infection in the United States: National Health and Nutrition Examination Survey 2001 through 2010," *Journal of Hepatology*, vol. 60, no. 4, pp. 691–698, 2014.

[62] dOnofrio, A., Ledzewicz, U., Maurer, H., and Schättler, H., "On optimal delivery of combination therapy for tumors," *Mathematical biosciences*, vol. 222, no. 1, pp. 13–26, 2009.

[63] Dowdle, W. R. and Cochi, S. L., "The principles and feasibility of disease eradication," *Vaccine*, vol. 29, pp. D70–D73, 2011.

[64] Drummond, M., Sculpher, M., Torrance, G., O'Brien, B., and Stoddart, G., *Methods for the economic evaluation of health care programmes*. Oxford University Press, 2005.

[65] Echevarria, D., Gutfraind, A., Boodram, B., Major, M., Del Valle, S., Cotler, S. J., and Dahari, H., "Mathematical modeling of hepatitis c prevalence reduction with antiviral treatment scale-up in persons who inject drugs in metropolitan chicago," *PloS one*, vol. 10, no. 8, p. e0135901, 2015.

[66] Edlin, B. R., "Perspective: test and treat this silent killer," *Nature*, vol. 474, no. 7350_supp, pp. s18–s19, 2011.

[67] Eichhorst, B. F., Busch, R., Stilgenbauer, S., Stauch, M., Bergmann, M. A., Ritgen, M., Kranzhfer, N., Rohrberg, R., Sling, U., and Burkhard, O., "First-line therapy with fludarabine compared with chlorambucil does not result in a major benefit for elderly patients with advanced chronic lymphocytic leukemia," *Blood*, vol. 114, no. 16, pp. 3382–3391, 2009.

[68] El-Serag, H., "Hepatocellular carcinoma," *New England Journal of Medicine*, vol. 365, no. 12, pp. 1118–1127, 2011.

[69] El-Serag, H. and Davila, J., "Surveillance for hepatocellular carcinoma: In whom and how?," *Therapeutic Advances in Gastroenterology*, vol. 4, no. 1, pp. 5–10, 2011.

[70] El-Serag, H., Siegel, A., Davila, J., Shaib, Y., Cayton-Woody, M., McBride, R., and McGlynn, K., "Treatment and outcomes of treating of hepatocellular carcinoma among medicare recipients in the United States: A population-based study," *Journal of hepatology*, vol. 44, no. 1, pp. 158–166, 2006.

[71] Elbasha, E., "Model for hepatitis c virus transmissions.," *Mathematical biosciences and engineering: MBE*, vol. 10, no. 4, pp. 1045–1065, 2013.

[72] Farooqui, M. Z., Valdez, J., Martyr, S., Aue, G., Saba, N., Niemann, C. U., Herman, S. E., Tian, X., Marti, G., and Soto, S., "Ibrutinib for previously untreated and relapsed or refractory chronic lymphocytic leukaemia with tp53 aberrations: a phase 2, single-arm trial," *The Lancet Oncology*, vol. 16, no. 2, pp. 169–176, 2015.

[73] Fattovich, G., Giustina, G., Degos, F., Tremolada, F., Diodati, G., Almasio, P., Nevens, F., Solinas, A., Mura, D., Brouwer, J., and others, "Morbidity and mortality in compensated cirrhosis type C: A retrospective follow-up study of 384 patients," *Gastroenterology*, vol. 112, no. 2, pp. 463–472, 1997.

[74] Fenwick, E., Claxton, K., and Sculpher, M., "Representing uncertainty: The role of cost-effectiveness acceptability curves," *Health economics*, vol. 10, no. 8, pp. 779–787, 2001.

[75] Ferguson, J., Tolley, K., Gilmour, L., and Priaulx, J., "Pcn79 health state preferences study mapping the change over the course of the disease process in chronic lymphocytic leukemia (cll)," *Value in Health*, vol. 11, no. 6, p. A485, 2008.

[76] Fischer, K., Cramer, P., Busch, R., Stilgenbauer, S., Bahlo, J., Schweighofer, C. D., Bttcher, S., Staib, P., Kiehl, M., and Eckart, M. J., "Bendamustine combined with rituximab in patients with relapsed and/or refractory chronic lymphocytic leukemia: a multicenter phase ii trial of the german chronic lymphocytic leukemia study group," *Journal of Clinical Oncology*, vol. 29, no. 26, pp. 3559–3566, 2011.

[77] Food and Drug Administration, "FDA approves new drug for chronic lymphocytic leukemia in patients with a specific chromosomal abnormality." `http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm495253.htm`. Accessed: 2016-10-15.

[78] Forner, A., Llovet, J., and Bruix, J., "Hepatocellular carcinoma," *The Lancet*, vol. 379, no. 9822, pp. 1245–1255, 2012.

[79] Forner, A., Reig, M., de Lope, C., and Bruix, J., "Current strategy for staging and treatment: The BCLC update and future prospects," in *Seminars in liver disease*, vol. 30, pp. 061–074, 2010.

[80] Forster, G. A. and Gilligan, C. A., "Optimizing the control of disease infestations at the landscape scale," *Proceedings of the National Academy of Sciences*, vol. 104, no. 12, pp. 4984–4989, 2007.

[81] Furman, R. R., Sharman, J. P., Coutre, S. E., Cheson, B. D., Pagel, J. M., Hillmen, P., Barrientos, J. C., Zelenetz, A. D., Kipps, T. J., and Flinn, I., "Idelalisib and rituximab in relapsed chronic lymphocytic leukemia," *New England Journal of Medicine*, vol. 370, no. 11, pp. 997–1007, 2014.

[82] Geering, H. P., *Optimal control with engineering applications*, vol. 113. Springer, 2007.

[83] Gerteis, J., Izrael, D., Deitz, D., LeRoy, L., Ricciardi, R., Miller, T., and Basu, J., "Multiple chronic conditions chartbook," 2014.

[84] Goede, V., Fischer, K., Busch, R., Engelke, A., Eichhorst, B., Wendtner, C. M., Chagorova, T., de la Serna, J., Dilhuydy, M.-S., and Illmer, T., "Obinutuzumab plus chlorambucil in patients with cll and coexisting conditions," *New England Journal of Medicine*, vol. 370, no. 12, pp. 1101–1110, 2014.

[85] Gold, M., Siegel, J., Russell, L., and Weinstein, M., *Cost-effectiveness in health and medicine*. 1996.

[86] Goldstein, D., Ahmad, B., Chen, Q., Ayer, T., Howard, D., Lipscomb, J., El-Rayes, B., and Flowers, C., "Cost-effectiveness analysis of regorafenib for metastatic colorectal cancer.," *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, vol. 33, no. 32, pp. 3727–3732, 2015.

[87] Goldstein, D. A., Chen, Q., Ayer, T., Howard, D. H., Lipscomb, J., El-Rayes, B. F., and Flowers, C. R., "First-and second-line bevacizumab in addition to chemotherapy for metastatic colorectal cancer: A united statesbased cost-effectiveness analysis," *Journal of Clinical Oncology*, vol. 33, no. 10, pp. 1112–1118,

2015.

[88] Gollogly, L. and others, *World health statistics 2009*. World Health Organization, 2009.

[89] Grebely, J., Conway, B., Raffa, J. D., Lai, C., Krajden, M., and Tyndall, M. W., "Hepatitis c virus reinfection in injection drug users," *Hepatology*, vol. 44, no. 5, pp. 1139–1145, 2006.

[90] Greenland, P. and Gaziano, J., "Selecting asymptomatic patients for coronary computed tomography or electrocardiographic exercise testing," *New England Journal of Medicine*, vol. 349, no. 5, pp. 465–473, 2003.

[91] Greten, T., Papendorf, F., Bleck, J., Kirchhoff, T., Wohlberedt, T., Kubicka, S., Klempnauer, J., Galanski, M., and Manns, M., "Survival rate in patients with hepatocellular carcinoma: A retrospective analysis of 389 patients," *British journal of cancer*, vol. 92, no. 10, pp. 1862–1868, 2005.

[92] Gribben, J. G. and O'Brien, S., "Update on therapy of chronic lymphocytic leukemia," *J Clin Oncol*, vol. 29, no. 5, pp. 544–50, 2011.

[93] Guadagnolo, B. A., Punglia, R. S., Kuntz, K. M., Mauch, P. M., and Ng, A. K., "Cost-effectiveness analysis of computerized tomography in the routine follow-up of patients after primary treatment for hodgkins disease," *Journal of clinical oncology*, vol. 24, no. 25, pp. 4116–4122, 2006.

[94] Hahn, J. A., Wylie, D., Dill, J., Sanchez, M. S., Lloyd-Smith, J. O., Page-Shafer, K., and Getz, W. M., "Potential impact of vaccination on the hepatitis c virus epidemic in injection drug users," *Epidemics*, vol. 1, no. 1, pp. 47–57, 2009.

[95] Hallek, M., Fischer, K., Fingerle-Rowson, G., Fink, A., Busch, R., Mayer, J., Hensel, M., Hopfinger, G., Hess, G., and Von Gruenhagen, U., "Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial," *The Lancet*, vol. 376, no. 9747, pp. 1164–1174, 2010.

[96] Hallek, M., "Chronic lymphocytic leukemia: 2015 update on diagnosis, risk stratification, and treatment," *American Journal of Hematology*, vol. 90, no. 5, pp. 446–460, 2015.

[97] Hanmer, J., Lawrence, W., Anderson, J., Kaplan, R., and Fryback, D., "Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores," *Medical Decision Making*, vol. 26, no. 4, pp. 391–400, 2006.

[98] Hansen, E. and Day, T., "Optimal control of epidemics with limited resources," *Journal of mathematical biology*, vol. 62, no. 3, pp. 423–451, 2011.

[99] Hatzakis, A., Chulanov, V., Gadano, A., Bergin, C., Ben-Ari, Z., Mossong, J., Schréter, I., Baatarkhuu, O., Acharya, S., Aho, I., and others, "The present and future disease burden of hepatitis c virus (HCV) infections with today's treatment paradigm–volume 2," *Journal of viral hepatitis*, vol. 22, no. s1, pp. 26–45, 2015.

[100] He, T., Li, K., Roberts, M. S., Spaulding, A. C., Ayer, T., Grefenstette, J. J., and Chhatwal, J., "Prevention of hepatitis c by screening and treatment in us prisons," *Annals of internal medicine*, vol. 164, no. 2, pp. 84–92, 2016.

[101] Heffernan, J., Smith, R., and Wahl, L., "Perspectives on the basic reproductive ratio," *Journal of the Royal Society Interface*, vol. 2, no. 4, pp. 281–293, 2005.

[102] Hethcote, H. W., "The mathematics of infectious diseases," *SIAM review*, vol. 42, no. 4, pp. 599–653, 2000.

[103] Himmelstein, D. U., Thorne, D., Warren, E., and Woolhandler, S., "Medical bankruptcy in the united states, 2007: results of a national study," *The American journal of medicine*, vol. 122, no. 8, pp. 741–746, 2009.

[104] Holtz-Eakin, D., "Prices for brand-name drugs under selected federal programs. washington, dc: Congressional budget office; 2005," 2011.

[105] Ikai, I., Arii, S., Kojiro, M., Ichida, T., Makuuchi, M., Matsuyama, Y., Nakanuma, Y., Okita, K., Omata, M., Takayasu, K., and others, "Reevaluation of prognostic factors for survival after liver resection in patients with hepatocellular carcinoma in a Japanese nationwide survey," *Cancer*, vol. 101, no. 4, pp. 796–802, 2004.

[106] Inadomi, J., Sampliner, R., Lagergren, J., Lieberman, D., Fendrick, A., and Vakil, N., "Screening and surveillance for Barrett esophagus in high-risk groups: A cost–utility analysis," *Annals of Internal Medicine*, vol. 138, no. 3, pp. 176–186, 2003.

[107] Jain, N. and O'Brien, S., "Initial treatment of cll: integrating biology and functional status," *Blood*, vol. 126, no. 4, pp. 463–70, 2015.

[108] Jain, N. and O'Brien, S., "Targeted therapies for cll: Practical issues with the changing treatment paradigm," *Blood Rev*, 2015.

[109] Jayachandran, D., Rundell, A. E., Hannemann, R. E., Vik, T. A., and Ramkrishna, D., "Optimal chemotherapy for leukemia: a model-based strategy for individualized treatment," *PloS one*, vol. 9, no. 10, p. e109623, 2014.

[110] Kabiri, M., Jazwinski, A., Roberts, M., Schaefer, A., and Chhatwal, J., "The changing burden of hepatitis C virus infection in the United States: Model-based predictions," *Annals of internal medicine*, vol. 161, no. 3, pp. 170–180, 2014.

[111] Kale, H. P. and Carroll, N. V., "Self-reported financial burden of cancer care and its effect on physical and mental health-related quality of life among us cancer survivors," *Cancer*, vol. 122, no. 8, pp. 283–289, 2016.

[112] Kallenberg, L. C. M., "Survey of linear programming for standard and nonstandard markovian control problems. part i: Theory," *Zeitschrift für Operations Research*, vol. 40, no. 1, pp. 1–42, 1994.

[113] Kansagara, D., Papak, J., Pasha, A., ONeil, M., Freeman, M., Relevo, R., Quiñones, A., Motuapuaka, M., and Jou, J., "Screening for hepatocellular carcinoma in chronic liver disease: A systematic review," *Annals of internal medicine*, vol. 161, no. 4, pp. 261–269, 2014.

[114] Kantarjian, H., Steensma, D., Sanjuan, J. R., Elshaug, A., and Light, D., "High cancer drug prices in the united states: reasons and proposed solutions," *Journal of Oncology Practice*, vol. 10, no. 4, pp. e208–e211, 2014.

[115] Keating, M. J., O'Brien, S., Albitar, M., Lerner, S., Plunkett, W., Giles, F., Andreeff, M., Cortes, J., Faderl, S., Thomas, D., Koller, C., Wierda,

W., Detry, M. A., Lynn, A., and Kantarjian, H., "Early results of a chemoimmunotherapy regimen of fludarabine, cyclophosphamide, and rituximab as initial therapy for chronic lymphocytic leukemia," *J Clin Oncol*, vol. 23, no. 18, pp. 4079–88, 2005.

[116] Khademi, A., Saure, D. R., Schaefer, A. J., Braithwaite, R. S., and Roberts, M. S., "The price of nonabandonment: Hiv in resource-limited settings," *Manufacturing & Service Operations Management*, vol. 17, no. 4, pp. 554–570, 2015.

[117] Khalil, H. K. and Grizzle, J., *Nonlinear systems*, vol. 3. Prentice hall New Jersey, 1996.

[118] Kirk, D. E., *Optimal control theory: an introduction*. Courier Corporation, 2012.

[119] Kirkizlar, E., Faissol, D., Griffin, P., and Swann, J., "Timing of testing and treatment for asymptomatic diseases," *Mathematical biosciences*, vol. 226, no. 1, pp. 28–37, 2010.

[120] Kok, S., Rutherford, A. R., Gustafson, R., Barrios, R., Montaner, J. S., Vasarhelyi, K., and others, "Optimizing an hiv testing program using a system dynamics model of the continuum of care," *Health care management science*, vol. 18, no. 3, pp. 334–362, 2015.

[121] Kongnakorn, T., Sterchele, J. A., Salvador, C. G., Getsios, D., and Mwamburi, M., "Economic implications of using bendamustine, alemtuzumab, or chlorambucil as a first-line therapy for chronic lymphocytic leukemia in the us: a cost-effectiveness analysis," *ClinicoEconomics and outcomes research: CEOR*, vol. 6, p. 141, 2014.

[122] Korobeinikov, A. and Wake, G. C., "Lyapunov functions and global stability for SIR, SIRS, and SIS epidemiological models," *Applied Mathematics Letters*, vol. 15, no. 8, pp. 955–960, 2002.

[123] Lafeuille, M.-H., Vekeman, F., Wang, S.-T., Kerrigan, M., Menditto, L., and Duh, M. S., "Lifetime costs to medicare of providing care to patients with chronic lymphocytic leukemia," *Leukemia & lymphoma*, vol. 53, no. 6, pp. 1146–1154, 2012.

[124] Lakshmikantham, V., Leela, S., and Martynyuk, A. A., *Stability Analysis of Nonlinear Systems*. Birkhäuser, 2015.

[125] Lang, K., Danchenko, N., Gondek, K., Shah, S., and Thompson, D., "The burden of illness associated with hepatocellular carcinoma in the United States," *Journal of hepatology*, vol. 50, no. 1, pp. 89–99, 2009.

[126] Larsson, S., Mantzoros, C., and Wolk, A., "Diabetes mellitus and risk of breast cancer: a meta-analysis," *International journal of cancer*, vol. 121, no. 4, pp. 856–862, 2007.

[127] Lasry, A., Zaric, G. S., and Carter, M. W., "Multi-level resource allocation for hiv prevention: A model for developing countries," *European Journal of Operational Research*, vol. 180, no. 2, pp. 786–799, 2007.

[128] Ledzewicz, U. and Schättler, H., "Antiangiogenic therapy in cancer treatment as an optimal control problem," *SIAM Journal on Control and Optimization*, vol. 46, no. 3, pp. 1052–1079, 2007.

[129] Ledzewicz, U. and Schättler, H., "On optimal singular controls for a general sir-model with vaccination and treatment," *Discrete and Continuous Dynamical Systems*, pp. 981–990, 2011.

[130] Lee, W., Dienstag, J., Lindsay, K., Lok, A., Bonkovsky, H., Shiffman, M., Everson, G., Di Bisceglie, A., Morgan, T., Ghany, M., and others, "Evolution of the HALT-C trial: Pegylated interferon as maintenance therapy for chronic hepatitis C in previous interferon nonresponders," *Controlled clinical trials*, vol. 25, no. 5, pp. 472–492, 2004.

[131] Lee, W. C., Lamas, G. A., Balu, S., Spalding, J., Wang, Q., and Pashos, C. L., "Direct treatment cost of atrial fibrillation in the elderly american population: a medicare perspective," *Journal of medical economics*, vol. 11, no. 2, pp. 281–298, 2008.

[132] Lencioni, R., Cioni, D., Crocetti, L., Franchini, C., Della Pina, C., Lera, J., and Bartolozzi, C., "Early-stage hepatocellular carcinoma in patients with cirrhosis: Long-term results of percutaneous image-guided radiofrequency ablation," *Radiology*, vol. 234, no. 3, pp. 961–967, 2005.

[133] Leoni, S., Piscaglia, F., Golfieri, R., Camaggi, V., Vidili, G., Pini, P., and Bolondi, L., "The impact of vascular and nonvascular findings on the noninvasive diagnosis of small hepatocellular carcinoma based on the EASL and AALSD criteria," *The American journal of gastroenterology*, vol. 105, no. 3, pp. 599–609, 2009.

[134] Levinson, D. R., "Medicaid drug price comparison: Average sales price to average wholesale price," *Office of the Inspector General. Department of Health and Human Serives*, 2005.

[135] Liberzon, D., *Calculus of variations and optimal control theory: a concise introduction*. Princeton University Press, 2012.

[136] Light, D. W. and Kantarjian, H., "Market spiral pricing of cancer drugs," *Cancer*, vol. 119, no. 22, pp. 3900–3902, 2013.

[137] Lin, O., Keeffe, E., Sanders, G., and Owens, D., "Cost-effectiveness of screening for hepatocellular carcinoma in patients with cirrhosis due to chronic hepatitis C," *Alimentary pharmacology & therapeutics*, vol. 19, no. 11, pp. 1159–1172, 2004.

[138] Linas, B. P., Barter, D. M., Leff, J. A., Assoumou, S. A., Salomon, J. A., Weinstein, M. C., Kim, A. Y., and Schackman, B. R., "The hepatitis c cascade of care: identifying priorities to improve clinical outcomes," *PloS one*, vol. 9, no. 5, p. e97317, 2014.

[139] Liu, S., Cipriano, L., Holodniy, M., Owens, D., and Goldhaber-Fiebert, J., "New protease inhibitors for the treatment of chronic hepatitis C," *Ann Intern Med*, vol. 156, pp. 279–290, 2012.

[140] Livraghi, T., Meloni, F., Di Stasi, M., Rolle, E., Solbiati, L., Tinelli, C., and Rossi, S., "Sustained complete response and complications rates after radiofrequency ablation of very early hepatocellular carcinoma in cirrhosis: Is resection still the treatment of choice?," *Hepatology*, vol. 47, no. 1, pp. 82–89, 2008.

[141] Llovet, J. and Bruix, J., "Systematic review of randomized trials for unresectable hepatocellular carcinoma: Chemoembolization improves survival," *Hepatology*, vol. 37, no. 2, pp. 429–442, 2003.

[142] Llovet, J., Di Bisceglie, A., Bruix, J., Kramer, B., Lencioni, R., Zhu, A., Sherman, M., Schwartz, M., Lotze, M., Talwalkar, J., and others, "Design and endpoints of clinical trials in hepatocellular carcinoma," *Journal of the National Cancer Institute*, vol. 100, no. 10, pp. 698–711, 2008.

[143] Llovet, J. and Ducreux, M., "EASL–EORTC clinical practice guidelines: Management of hepatocellular carcinoma," *Journal of Hepatology*, vol. 56, no. 4, pp. 908–943, 2012.

[144] Lok, A., Seeff, L., Morgan, T., Di Bisceglie, A., Sterling, R., Curto, T., Everson, G., Lindsay, K., Lee, W., Bonkovsky, H., and others, "Incidence of hepatocellular carcinoma and associated risk factors in hepatitis C-related advanced liver disease," *Gastroenterology*, vol. 136, no. 1, pp. 138–148, 2009.

[145] Ly, K., Xing, J., Klevens, R., Jiles, R., Ward, J., and Holmberg, S., "The increasing burden of mortality from viral hepatitis in the United States between 1999 and 2007," *Annals of Internal Medicine*, vol. 156, no. 4, pp. 271–278, 2012.

[146] Maddocks, K. J., Ruppert, A. S., Lozanski, G., Heerema, N. A., Zhao, W., Abruzzo, L., Lozanski, A., Davis, M., Gordon, A., and Smith, L. L., "Etiology of ibrutinib therapy discontinuation and outcomes in patients with chronic lymphocytic leukemia," *JAMA Oncology*, vol. 1, no. 1, pp. 80–87, 2015.

[147] Mahon, F.-X., Réa, D., Guilhot, J., Guilhot, F., Huguet, F., Nicolini, F., Legros, L., Charbonnier, A., Guerci, A., Varet, B., and others, "Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre stop imatinib (stim) trial," *The lancet oncology*, vol. 11, no. 11, pp. 1029–1035, 2010.

[148] Maillart, L., Ivy, J., Ransom, S., and Diehl, K., "Assessing dynamic breast cancer screening policies," *Operations Research*, vol. 56, no. 6, pp. 1411–1427, 2008.

[149] Malvankar-Mehta, M. S. and Xie, B., "Optimal incentives for allocating hiv/aids prevention resources among multiple populations," *Health care management science*, vol. 15, no. 4, pp. 327–338, 2012.

[150] Mariotto, A. B., Yabroff, K. R., Shao, Y., Feuer, E. J., and Brown, M. L., "Projections of the cost of cancer care in the united states: 2010–2020," *Journal of the National Cancer Institute*, 2011.

[151] Marsh, K., Xu, P., Orfanos, P., Gordon, J., and Griebsch, I., "Model-based cost-effectiveness analyses for the treatment of chronic lymphocytic leukaemia: A review of methods to model disease outcomes and estimate utility," *PharmacoEconomics*, vol. 32, no. 10, pp. 981–993, 2014.

[152] Martcheva, M., *Introduction to Mathematical Epidemiology*, vol. 61. Springer, 2015.

[153] Martin, N. K., Hickman, M., Hutchinson, S. J., Goldberg, D. J., and Vickerman, P., "Combination interventions to prevent HCV transmission among people who inject drugs: modeling the impact of antiviral treatment, needle and syringe programs, and opiate substitution therapy," *Clinical Infectious Diseases*, vol. 57, no. suppl 2, pp. S39–S45, 2013.

[154] Martin, N. K., Pitcher, A. B., Vickerman, P., Vassall, A., and Hickman, M., "Optimal control of hepatitis c antiviral treatment programme delivery for prevention amongst a population of injecting drug users," *PLoS One*, vol. 6, no. 8, p. e22309, 2011.

[155] Martin, N. K., Vickerman, P., Foster, G. R., Hutchinson, S. J., Goldberg, D. J., and Hickman, M., "Can antiviral therapy for hepatitis c reduce the prevalence of HCV among injecting drug user populations? a modeling analysis of its prevention utility," *Journal of hepatology*, vol. 54, no. 6, pp. 1137–1144, 2011.

[156] Martin, N. K., Vickerman, P., Grebely, J., Hellard, M., Hutchinson, S. J., Lima, V. D., Foster, G. R., Dillon, J. F., Goldberg, D. J., Dore, G. J., and others, "Hepatitis c virus treatment for prevention among people who inject drugs: Modeling treatment scale-up in the age of direct-acting antivirals," *Hepatology*, vol. 58, no. 5, pp. 1598–1609, 2013.

[157] Martin, N. K., Vickerman, P., and Hickman, M., "Mathematical modelling of hepatitis c treatment for injecting drug users," *Journal of Theoretical Biology*, vol. 274, no. 1, pp. 58–66, 2011.

[158] Maurer, H., "Numerical solution of singular control problems using multiple shooting techniques," *Journal of Optimization Theory and Applications*, vol. 18, no. 2, pp. 235–257, 1976.

[159] Mbah, M. L. N. and Gilligan, C. A., "Optimization of control strategies for epidemics in heterogeneous populations with symmetric and asymmetric transmission," *Journal of Theoretical Biology*, vol. 262, no. 4, pp. 757–763, 2010.

[160] Mbah, M. L. N. and Gilligan, C. A., "Resource allocation for epidemic control in metapopulations," *PLoS One*, vol. 6, no. 9, p. e24577, 2011.

[161] McAdam-Marx, C., McGarry, L., Hane, C., Biskupiak, J., Deniz, B., and Brixner, D., "All-cause and incremental per patient per year cost associated with chronic hepatitis C virus and associated liver complications in the United States: A managed care perspective," *Journal of Managed Care Pharmacy*, vol. 17, no. 7, p. 531, 2011.

[162] McCluskey, C., "Lyapunov functions for tuberculosis models with fast and slow progression.," *Mathematical biosciences and engineering: MBE*, vol. 3, no. 4, pp. 603–614, 2006.

[163] Meropol, N. J., Schrag, D., Smith, T. J., Mulvey, T. M., Langdon, R. M., Blum, D., Ubel, P. A., and Schnipper, L. E., "American society of clinical oncology guidance statement: the cost of cancer care," *Journal of Clinical Oncology*, vol. 27, no. 23, pp. 3868–3874, 2009.

[164] Miller, D. K. and Homan, S. M., "Determining transition probabilities confusion and suggestions," *Medical Decision Making*, vol. 14, no. 1, pp. 52–58, 1994.

[165] Mohd Hanafiah, K., Groeger, J., Flaxman, A. D., and Wiersma, S. T., "Global epidemiology of hepatitis c virus infection: New estimates of age-specific antibody to HCV seroprevalence," *Hepatology*, vol. 57, no. 4, pp. 1333–1342, 2013.

[166] Neumann, P., Cohen, J., and Weinstein, M., "Updating cost-effectivenessthe curious resilience of the $50,000-per-QALY threshold," *New England Journal of Medicine*, vol. 371, no. 9, pp. 796–797, 2014.

[167] Neumann, P. and Weinstein, M., "Legislating against use of cost-effectiveness information," *New England Journal of Medicine*, vol. 363, no. 16, pp. 1495–1497, 2010.

[168] OPTN, "2012 annual report of the US Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients: Transplant data," 2012.

[169] Organization, W. H. and others, "Global health sector strategy on viral hepatitis 2016-2021. towards ending viral hepatitis," 2016.

[170] Ortman, J. M., Velkoff, V. A., and Hogan, H., "An aging nation: the older

population in the united states," *Washington, DC: US Census Bureau*, pp. 25–1140, 2014.

[171] Parikh, S. A., Rabe, K. G., Kay, N. E., Call, T. G., Ding, W., Schwager, S. M., Bowen, D. A., Conte, M., Jelinek, D. F., Slager, S. L., and Shanafelt, T. D., "Chronic lymphocytic leukemia in young ($\leq$ 55 years) patients: A comprehensive analysis of prognostic factors and outcomes," *Haematologica*, vol. 99, no. 1, pp. 140–7, 2014.

[172] Patterson, M. A. and Rao, A. V., "GPOPS-II: A MATLAB software for solving multiple-phase optimal control problems using hp-adaptive Gaussian quadrature collocation methods and sparse nonlinear programming," *ACM Transactions on Mathematical Software (TOMS)*, vol. 41, no. 1, p. 1, 2014.

[173] Planas, R., Ballesté, B., Antonio Álvarez, M., Rivera, M., Montoliu, S., Anton Galeras, J., Santos, J., Coll, S., Maria Morillas, R., and Solà, R., "Natural history of decompensated hepatitis C virus-related cirrhosis: A study of 200 patients," *Journal of hepatology*, vol. 40, no. 5, pp. 823–830, 2004.

[174] Poustchi, H., Farrell, G., Strasser, S., Lee, A., McCaughan, G., and George, J., "Feasibility of conducting a randomized control trial for liver cancer screening: Is a randomized controlled trial for liver cancer screening feasible or still needed?," *Hepatology*, vol. 54, no. 6, pp. 1998–2004, 2011.

[175] Puterman, M., *Markov decision processes: Discrete stochastic dynamic programming*, vol. 414. John Wiley & Sons, 2009.

[176] Ramsey, S., Blough, D., Kirchhoff, A., Kreizenbeck, K., Fedorenko, C., Snell, K., Newcomb, P., Hollingworth, W., and Overstreet, K., "Washington state cancer patients found to be at greater risk for bankruptcy than people without a cancer diagnosis," *Health affairs*, vol. 32, no. 6, pp. 1143–1152, 2013.

[177] Ramsey, S. D., Bansal, A., Fedorenko, C. R., Blough, D. K., Overstreet, K. A., Shankaran, V., and Newcomb, P., "Financial insolvency as a risk factor for early mortality among patients with cancer," *Journal of Clinical Oncology*, vol. 34, no. 9, pp. 980–986, 2016.

[178] Rao, A. V., "A survey of numerical methods for optimal control," *Advances in the Astronautical Sciences*, vol. 135, no. 1, pp. 497–528, 2009.

[179] Ratcliffe, J., Longworth, L., Young, T., Bryan, S., Burroughs, A., and Buxton, M., "Assessing health-related quality of life pre–and post–liver transplantation: A prospective multicenter study," *Liver transplantation*, vol. 8, no. 3, pp. 263–270, 2002.

[180] Rauner, M., Gutjahr, W., Heidenberger, K., Wagner, J., and Pasia, J., "Dynamic policy modeling for chronic diseases: Metaheuristic-based identification of pareto-optimal screening strategies," *Operations research*, vol. 58, no. 5, pp. 1269–1286, 2010.

[181] Razavi, H., ElKhoury, A., Elbasha, E., Estes, C., Pasini, K., Poynard, T., and Kumar, R., "Chronic hepatitis C virus (HCV) disease burden and cost in the United States," *Hepatology*, vol. 57, no. 6, pp. 2164–2170, 2013.

[182] Ronot, M. and Vilgrain, V., "Hepatocellular carcinoma: Diagnostic criteria by imaging techniques," *Best Practice & Research Clinical Gastroenterology*, vol. 28, no. 5, pp. 795–812, 2014.

[183] Rowthorn, B. R. and Toxvaerd, F., "The optimal control of infectious diseases via prevention and treatment," 2012.

[184] Rowthorn, R. E., Laxminarayan, R., and Gilligan, C. A., "Optimal control of epidemics in metapopulations," *Journal of the Royal Society Interface*, vol. 6, no. 41, pp. 1135–1144, 2009.

[185] Ryerson, A. B., Eheman, C. R., Altekruse, S. F., Ward, J. W., Jemal, A., Sherman, R. L., Henley, S. J., Holtzman, D., Lake, A., Noone, A.-M., and others, "Annual report to the nation on the status of cancer, 1975-2012, featuring the increasing incidence of liver cancer," *Cancer*, vol. 122, no. 9, pp. 1312–1337, 2016.

[186] Saltz, L. B., "Perspectives on cost and value in cancer care," *JAMA oncology*, vol. 2, no. 1, pp. 19–21, 2016.

[187] Sandikçi, B., Maillart, L., Schaefer, A., Alagoz, O., and Roberts, M., "Estimating the patient's price of privacy in liver transplantation," *Operations Research*, vol. 56, no. 6, pp. 1393–1410, 2008.

[188] Sangiovanni, A., Manini, M., Iavarone, M., Romeo, R., Forzenigo, L., Fraquelli, M., Massironi, S., Della Corte, C., Ronchi, G., Rumi, M., and others, "The diagnostic and economic impact of contrast imaging techniques in the diagnosis of small hepatocellular carcinoma in cirrhosis," *Gut*, vol. 59, no. 5, pp. 638–644, 2010.

[189] Saving, C., "Why we should be willing to pay for hepatitis c treatment," *Clinical Gastroenterology and Hepatology*, vol. 13, pp. 1711–1713, 2015.

[190] Seierstad, A. and Sydsaeter, K., *Optimal control theory with economic applications*. Elsevier North-Holland, Inc., 1986.

[191] Sethi, S. and Thompson, G., *Optimal Control Theory: Applications to Management Science and Economics*. Springer, 2000.

[192] Shanafelt, T., Borah, B., Finnes, H., Chaffee, K., Ding, W., Leis, J., Chanan-Khan, A., Parikh, S., Slager, S., Kay, N., and others, "Impact of ibrutinib and idelalisib on the pharmaceutical cost of treating chronic lymphocytic leukemia at the individual and societal levels.," *Journal of oncology practice/American Society of Clinical Oncology*, vol. 11, no. 3, pp. 252–258, 2015.

[193] Shih, V., ten Ham, R. M., Bui, C. T., Tran, D. N., Ting, J., and Wilson, L., "Targeted therapies compared to dacarbazine for treatment of braf v600e metastatic melanoma: A cost-effectiveness analysis," *Journal of skin cancer*, vol. 2015, 2015.

[194] Shivkumar, S., Peeling, R., Jafari, Y., Joseph, L., and Pai, N. P., "Accuracy of rapid and point-of-care screening tests for hepatitis c: a systematic review and meta-analysis," *Annals of internal medicine*, vol. 157, no. 8, pp. 558–566, 2012.

[195] Shulgin, B., Stone, L., and Agur, Z., "Pulse vaccination strategy in the sir epidemic model," *Bulletin of mathematical biology*, vol. 60, no. 6, pp. 1123–1148, 1998.

[196] Siegel, R., Ward, E., Brawley, O., and Jemal, A., "Cancer statistics, 2011," *CA: A Cancer Journal for Clinicians*, 2011.

[197] Siegel, R., Ma, J., Zou, Z., and Jemal, A., "Cancer statistics, 2014," *CA: A Cancer Journal for Clinicians*, vol. 64, no. 1, pp. 9–29, 2014.

[198] Siegel, R., Naishadham, D., and Jemal, A., "Cancer statistics, 2012," *CA: A cancer journal for clinicians*, vol. 62, no. 1, pp. 10–29, 2012.

[199] Siegel, R., Naishadham, D., and Jemal, A., "Cancer statistics, 2013," *CA: A cancer journal for clinicians*, vol. 63, no. 1, pp. 11–30, 2013.

[200] Siegel, R. L., Miller, K. D., and Jemal, A., "Cancer statistics, 2015," *CA: A cancer journal for clinicians*, vol. 65, no. 1, pp. 5–29, 2015.

[201] Singal, A., Nehra, M., Adams-Huet, B., Yopp, A., Tiro, J., Marrero, J., Lok, A., and Lee, W., "Detection of hepatocellular carcinoma at advanced stages among patients in the HALT-C trial: Where did surveillance fail?," *The American journal of gastroenterology*, vol. 108, no. 3, pp. 425–432, 2013.

[202] Singal, A., Pillai, A., and Tiro, J., "Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: A meta-analysis," *PLoS medicine*, vol. 11, no. 4, p. e1001624, 2014.

[203] Singal, A., Volk, M., Waljee, A., Salgia, R., Higgins, P., Rogers, M., and Marrero, J., "Meta-analysis: Surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis," *Alimentary pharmacology & therapeutics*, vol. 30, no. 1, pp. 37–47, 2009.

[204] Singal, A., Yopp, A., Skinner, C., Packer, M., Lee, W., and Tiro, J., "Utilization of hepatocellular carcinoma surveillance among American patients: A systematic review," *Journal of general internal medicine*, vol. 27, no. 7, pp. 861–867, 2012.

[205] Sonnenberg, F. and Beck, J., "Markov models in medical decision making: A practical guide," *Medical decision making*, vol. 13, no. 4, pp. 322–338, 1993.

[206] Sood, A., Sarin, S. K., Midha, V., Hissar, S., Sood, N., Bansal, P., and Bansal, M., "Prevalence of hepatitis c virus in a selected geographical area of northern india: a population based survey," *Indian Journal of Gastroenterology*, vol. 31, no. 5, pp. 232–236, 2012.

[207] Steel, J., Chopra, K., Olek, M., and Carr, B., "Health-related quality of life: Hepatocellular carcinoma, chronic liver disease, and the general population," *Quality of Life Research*, vol. 16, no. 2, pp. 203–215, 2007.

[208] Stinnett, A. and Mullahy, J., "Net health benefits a new framework for the analysis of uncertainty in cost-effectiveness analysis," *Medical Decision Making*, vol. 18, no. 2, pp. S68–S80, 1998.

[209] Stravitz, R., Heuman, D., Chand, N., Sterling, R., Shiffman, M., Luketic, V., Sanyal, A., Habib, A., Mihas, A., Giles, H., and others, "Surveillance for hepatocellular carcinoma in patients with cirrhosis improves outcome," *The American journal of medicine*, vol. 121, no. 2, pp. 119–126, 2008.

[210] Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov). Prevalence database, "US Estimated 36-Year L-D Prevalence Counts on 1/1/2011. National Cancer Institute, DCCPS, Surveillance Research Program, Data Modeling Branch, released April 2014, based on the November 2013 SEER data submission.."

[211] Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database, "Incidence - SEER 9 Regs Research Data, Nov 2014 Sub (1973-2012) (Katrina/Rita Population Adjustment) - Linked To County Attributes - Total U.S., 1969-2013 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2015, based on the November 2014 submission.."

[212] Tam, C. S., O'Brien, S., Wierda, W., Kantarjian, H., Wen, S., Do, K. A., Thomas, D. A., Cortes, J., Lerner, S., and Keating, M. J., "Long-term results of the fludarabine, cyclophosphamide, and rituximab regimen as initial therapy of chronic lymphocytic leukemia," *Blood*, vol. 112, no. 4, pp. 975–80, 2008.

[213] The Kaiser Family Foundation, "Health research & educational trust: 2014 employer health benefits survey," report.

[214] The World Bank, "World bank open data - death rate, crude (per 1,000 people)." `http://data.worldbank.org/indicator/SP.DYN.CDRT.IN`, 2016.

[215] The World Health Organization, "Development of the global health sector strategy on viral hepatitis, 2016-2021." `http://www.who.int/hiv/mediacentre/news/hepstrategy2016-21/en/`, 2016.

[216] The World Health Organization, "Viral hepatitis c treatment in india." `http://www.searo.who.int/india/mediacentre/events/2016/viral_hepatitis_c_treatment_in_india.pdf?ua=1`, 2016.

[217] Thein, H., Yi, Q., Dore, G., and Krahn, M., "Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: A meta-analysis and meta-regression," *Hepatology*, vol. 48, no. 2, pp. 418–431, 2008.

[218] Thompson, G. L., *Optimal Control Theory: Applications to Management Science and Economics*. Springer, 2006.

[219] Thompson, P. A., Tam, C. S., O'Brien, S. M., Wierda, W. G., Stingo, F., Plunkett, W., Smith, S. C., Kantarjian, H. M., Freireich, E. J., and Keating, M. J., "Fludarabine, cyclophosphamide and rituximab achieves long-term disease-free survival in ighv-mutated chronic lymphocytic leukemia," *Blood*, 2015.

[220] Toro, A., Pulvirenti, E., Palermo, F., and Di Carlo, I., "Health-related quality of life in patients with hepatocellular carcinoma after hepatic resection, transcatheter arterial chemoembolization, radiofrequency ablation or no treatment," *Surgical Oncology*, vol. 21, no. 1, pp. e23–e30, 2012.

[221] Tumeh, J. W., Moore, S. G., Shapiro, R., and Flowers, C. R., "Practical approach for using medicare data to estimate costs for cost-effectiveness analysis," 2005.

[222] Ubel, P. A., Abernethy, A. P., and Zafar, S. Y., "Full disclosureout-of-pocket costs as side effects," *New England Journal of Medicine*, vol. 369, no. 16, pp. 1484–1486, 2013.

[223] U.S. Census Bureau, Population Division, "Annual Estimates of the Resident Population for Selected Age Groups by Sex for the United States, States, Counties, and Puerto Rico Commonwealth and Municipios: April 1, 2010 to July 1, 2014. Release Date: June 2015."

[224] U.S. Census Bureau, Population Division, "Table 12-C. Projections of the Population by Age and Sex for the United States: 2010 to 2050 Constant Net International Migration Series (NP2009-T12-C). Release Date: December 16, 2009."

[225] Van den Driessche, P. and Watmough, J., "Further notes on the basic reproduction number," in *Mathematical Epidemiology*, pp. 159–178, Springer, 2008.

[226] Vargas-De-León, C., "On the global stability of sis, sir and sirs epidemic models with standard incidence," *Chaos, Solitons & Fractals*, vol. 44, no. 12, pp. 1106–1110, 2011.

[227] Vickerman, P., Martin, N., Turner, K., and Hickman, M., "Can needle and syringe programmes and opiate substitution therapy achieve substantial reductions in hepatitis c virus prevalence? model projections for different epidemic settings," *Addiction*, vol. 107, no. 11, pp. 1984–1995, 2012.

[228] Wedemeyer, H., Dore, G., and Ward, J., "Estimates on HCV disease burden worldwide–filling the gaps," *Journal of viral hepatitis*, vol. 22, no. s1, pp. 1–5, 2015.

[229] Wierda, W. G., O'Brien, S., Wang, X., Faderl, S., Ferrajoli, A., Do, K. A., Garcia-Manero, G., Cortes, J., Thomas, D., Koller, C. A., Burger, J. A., Lerner, S., Schlette, E., Abruzzo, L., Kantarjian, H. M., and Keating, M. J., "Multivariable model for time to first treatment in patients with chronic lymphocytic leukemia," *J Clin Oncol*, vol. 29, no. 31, pp. 4088–95, 2011.

[230] Wierda, W., OBrien, S., Wen, S., Faderl, S., Garcia-Manero, G., Thomas, D., Do, K.-A., Cortes, J., Koller, C., and Beran, M., "Chemoimmunotherapy with fludarabine, cyclophosphamide, and rituximab for relapsed and refractory chronic lymphocytic leukemia," *Journal of Clinical Oncology*, vol. 23, no. 18, pp. 4070–4078, 2005.

[231] Wierda, W. G., Kipps, T. J., Mayer, J., Stilgenbauer, S., Williams, C. D., Hellmann, A., Robak, T., Furman, R. R., Hillmen, P., and Trneny, M., "Ofatumumab as single-agent cd20 immunotherapy in fludarabine-refractory chronic lymphocytic leukemia," *Journal of Clinical Oncology*, vol. 28, no. 10, pp. 1749–1755, 2010.

[232] Wolfe, R., Roys, E., and Merion, R., "Trends in organ donation and transplantation in the United States, 1999–2008," *American Journal of Transplantation*, vol. 10, no. 4p2, pp. 961–972, 2010.

[233] Wu, S.-Y. and Green, A., "Projection of chronic illness prevalence and cost inflation," *Santa Monica, CA: RAND Health*, p. 2000, 2000.

[234] Xu, J., Murphy, S. L., Kochanek, K. D., and Bastian, B. A., "Deaths: final data for 2013.," *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, vol. 64, no. 2, pp. 1–118, 2016.

[235] Yabroff, K. R., Dowling, E. C., Guy, G. P., Banegas, M. P., Davidoff, A., Han, X., Virgo, K. S., McNeel, T. S., Chawla, N., Blanch-Hartigan, D., Kent, E. E., Li, C., Rodriguez, J. L., de Moor, J. S., Zheng, Z., Jemal, A., and Ekwueme, D. U., "Financial hardship associated with cancer in the united states: Findings from a population-based sample of adult cancer survivors," *Journal of Clinical Oncology*, vol. 34, no. 3, pp. 259–267, 2016.

[236] Yang, J. and Kim, W., "Surveillance for hepatocellular carcinoma in patients with cirrhosis," *Clinical Gastroenterology and Hepatology*, vol. 10, no. 1, pp. 16–21, 2012.

[237] Yang, Y., Goldhaber-Fiebert, J., and Wein, L., "Analyzing screening policies for childhood obesity," *Management science*, vol. 59, no. 4, pp. 782–795, 2013.

[238] Younossi, Z., Park, H., Gordon, S., Ferguson, J., Ahmed, A., Dieterich, D., and Saab, S., "Real-world outcomes of ledipasvir/sofosbuvir in treatment-naive patients with hepatitis c.," *The American journal of managed care*, vol. 22, no. 6 Spec No., pp. SP205–11, 2016.

[239] Zaman, G., Kang, Y. H., and Jung, I. H., "Optimal treatment of an sir epidemic model with time delay," *BioSystems*, vol. 98, no. 1, pp. 43–50, 2009.

[240] Zaric, G. S. and Brandeau, M. L., "Resource allocation for epidemic control over short time horizons," *Mathematical Biosciences*, vol. 171, no. 1, pp. 33–58, 2001.

[241] Zaric, G. S. and Brandeau, M. L., "Dynamic resource allocation for epidemic control in multiple populations," *Mathematical Medicine and Biology*, vol. 19, no. 4, pp. 235–255, 2002.

[242] Zenios, S. A. and Fuloria, P. C., "Managing the delivery of dialysis therapy: A multiclass fluid model analysis," *Management Science*, vol. 46, no. 10, pp. 1317–1336, 2000.

[243] Zhang, J., Denton, B., Balasubramanian, H., Shah, N., and Inman, B., "Optimization of prostate biopsy referral decisions," *Manufacturing & Service Operations Management*, vol. 14, no. 4, pp. 529–547, 2012.